

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 02:27:36 ; Search time 5105.53 Seconds
(without alignments)
5033.405 Million cell updates/sec

Title: US-09-623-514A-1
Perfect score: 1904
Sequence: 1 attcttagcttccttcctc.....ctatggtgaagcgagcgcg 1904

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 13736207 segs, 6740477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : EST.*
1: em_estda.*
2: em_esthum.*
3: em_estlin.*
4: em_estnu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_estc.*
9: gb_est1.*
10: gb_est2.*
11: gb_hic.*
12: gb_gss.*
13: em_gss_hum.*
14: em_gss_inv.*
15: em_gss_pln.*
16: em_gss_vit.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	405.8	21.3	729	10	BI422326 EST52992
2	361	19.0	399	10	BE523720 MAOH1STM
3	336.4	17.7	629	9	AA042298 24635 CD4
4	320.4	16.8	549	9	AI992543 701558181
5	314.2	15.8	539	10	BI422212 EST52878
6	300.4	15.8	605	12	AO958057
7	299.4	15.7	855	9	AM349274
8	261	13.7	261	10	BE525637
9	252.8	13.3	260	10	BE525720
10	249.6	13.1	561	9	AM035727
11	226.8	11.9	626	9	AM035727
12	220	11.6	663	12	AO958056
13	216	11.3	862	10	BM321213
14	204.6	10.7	572	10	BM309699
15	189.2	9.9	422	10	BF199515
16	187.6	9.9	515	9	AV925760
17	184	9.7	682	10	BE247899

18	183.6	9.6	763	10	BI308446
19	179.6	9.4	577	9	AM586836
20	172.6	9.1	634	10	BF634363
21	162.6	8.5	617	9	AV930897
22	161.4	8.5	531	10	BE356202
23	159	8.4	655	9	AM775077
24	156.4	8.2	574	10	BE003479
25	148.6	7.8	740	10	BI924307
26	139.6	7.3	565	10	BG078967
27	138	7.2	1064	10	BM476315
28	137.8	7.2	706	10	BM348355
29	136.8	7.2	528	9	AL381190
30	136.2	7.2	267	10	BF632251
31	132.4	7.0	746	12	BM466416
32	131.4	6.9	832	9	AM342534
33	128.4	6.7	701	10	BM334936
34	127.4	6.7	700	10	BM341343
35	126.8	6.7	433	9	AI441040
36	126.2	6.6	531	9	AI414161
37	123.4	6.5	1000	9	AF122129
38	121.8	6.4	681	10	BM267747
39	121.8	6.4	682	10	BM075161
40	119.4	6.3	670	10	BM338485
41	119.4	6.3	672	10	BM334036
42	119.4	6.3	819	10	BI225365
43	118.2	6.2	667	9	AU168524
44	117	6.1	857	9	AL520752
45	115.4	6.1	664	10	BM266971

ALIGNMENTS

RESULT 1
BI422326 729 bp mRNA linear EST 16-AUG-2001
LOCUS EST52992 tomato callus, TAMU Lycopersicon esculentum cDNA clone
DEFINITION CLE69N10 5' end, mRNA sequence.
ACCESSION BI422326
VERSION BI422326.1 GI:15196624
KEYWORDS EST.
SOURCE Lycopersicon
ORGANISM Lycopersicon esculentum
Eukaryote: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; easterids I; Solanales; Solanales; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 729)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.,
Liang,F., Upton,J., Craven,M.B., Bowman,C.L., Alm,S., Koning,
C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato callus tissue
Unpublished (1999)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>

TITLE JOURNAL
COMMENT
FEATURES
source

1..729
/organism="Lycopersicon esculentum"
/cultivar="T496"
/db_xref="taxon:4081"
/clone="CLE69N10"
/clone.lib="tomato callus, TAMU"
/tissue_type="callus"
/dev_stage="25-40 days old"
/lab_host="XLI Blue MR"
/note="Vector: pBluescript SK(-); Site1: EcoRI; Site2:
XhoI; supplier: Giovannoni laboratory; clec - Cytiledons
of seedlings 7-10 days post-germination were excised, cut
at both ends and placed on MS medium with no selection.
Mixed callus was harvested at 25 and 40 days and included

Aradiopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

RESULT	3	AA042298	639 bp	linear	FST 19-sep-1997
OCUS		AA042298			

OY 1541 ttccaggtccttggcttcttcacaaactacacagaaagtttg---ctaacgg 1597
 Db 484 ttccaggtccttggcttcttcacaaactacacagaaagtttg---ctaacgg 1597
 OY 1598 ttgggaacatgatcttcttggcttcttcacaaactacacagaaagtttg---ctaacgg 425
 Db 424 ttgggaacatgatcttcttggcttcttcacaaactacacagaaagtttg---ctaacgg 425
 OY 1658 ttataccagacaccttgatgaacaaagatcgatgcatgaacaaactatgctgactgc 1657
 Db 364 ttataccagacaccttgatgaacaaagatcgatgcatgaacaaactatgctgactgc 365
 RESULT 8
 BE525637
 LOCUS 261 bp mRNA linear EST 19-MAR-2001
 DEFINITION M63K05STM Arabidopsis developing seed Arabidopsis thaliana cDNA
 ACCESSION BE525637
 VERSION BE525637.1 GI:9783615
 KEYWORDS EST
 SOURCE
 ORGANISM
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de
 Illarduya, O., Jaworski, J.G., Ohlrogge, J. and Benning, C.
 A new set of Arabidopsis expressed sequence tags from developing
 seeds. The metabolic pathway from carbohydrates to seed oil
 20567808
 CONTACT: Benning, C
 Dept. of Biochemistry & Molecular Biology
 Michigan State University
 224 Biochemistry, Michigan State University, East Lansing, MI 48824
 Tel: 517 355 1609
 Fax: 517 353 9334
 Email: benning@msu.edu
 Clones were originally prepared at Michigan State University.
 Arabidopsis Biological Resource Center, The Ohio State University,
 309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210
 USA, FAX: 6142920603 TEL: 6142929371.
 Location/Qualifiers
 1..261
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone_lib="600034122R1"
 /clone_1lb="Arabidopsis developing seed"
 /tissue_type="seed"
 /dev_stage="5-13 days after flowering"
 /lab_host="E.coli"
 /note="Organ: Developing seed; Vector: pluescript SK-;
 Site_1: EcoRI; Site_2: XhoI"
 BASE COUNT 69 a 55 c 71 g 66 t
 ORIGIN
 Query Match
 Best Local Similarity 13.7%; Score 261; DB 10; Length 261;
 Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 411 aagagcgcgaggaaggaaggaagcgagatgcatgtagtatacgaacgtcgt 470
 Db 1 AAGAGCGCGGAGGAAGGAAGGAAGCGGATGCTACGTTTACGTAATCGACCGT 470
 OY 471 tccagctcatcgagagcgagagagtcacatgctcgcgacgaatctcaacaag 530
 Db 61 tccagctcatcgagagcgagagagtcacatgctcgcgacgaatctcaacaag 530
 61 tccagctcatcgagagcgagagagtcacatgctcgcgacgaatctcaacaag 120

OY 531 catgcccgaatattacactcgtgtagtagttcttattgctgtaaacgtaacat 590
 Db 121 CCATCCCGGATATTATACACCTCGTGTAGTGTCTTATGCTGTAAACAGTACAT 590
 OY 591 catgaaaatcttaagaatagtggttgatgatacgaacgaatcttctgttagtcaag 180
 Db 181 CATGAAAATCTTATGAAGTATGCTGTGTGATCAACGCAATTTCTGTGTAGTCAAG 240
 OY 651 atcgcgcgagatggcgccgt 671
 Db 241 ATGCGCGGAGATGGCGCGCT 261
 RESULT 9
 BE525720
 LOCUS 260 bp mRNA linear EST 19-MAR-2001
 DEFINITION M63K06STM Arabidopsis developing seed Arabidopsis thaliana cDNA
 ACCESSION BE525720
 VERSION BE525720.1 GI:9783698
 KEYWORDS EST
 SOURCE
 ORGANISM
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de
 Illarduya, O., Jaworski, J.G., Ohlrogge, J. and Benning, C.
 A new set of Arabidopsis expressed sequence tags from developing
 seeds. The metabolic pathway from carbohydrates to seed oil
 20567808
 CONTACT: Benning, C
 Dept. of Biochemistry & Molecular Biology
 Michigan State University
 224 Biochemistry, Michigan State University, East Lansing, MI 48824
 Tel: 517 355 1609
 Fax: 517 353 9334
 Email: benning@msu.edu
 Clones were originally prepared at Michigan State University.
 Arabidopsis Biological Resource Center, The Ohio State University,
 309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210
 USA, FAX: 6142920603 TEL: 6142929371.
 Location/Qualifiers
 1..260
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone_lib="600034222R1"
 /clone_1lb="Arabidopsis developing seed"
 /tissue_type="seed"
 /dev_stage="5-13 days after flowering"
 /lab_host="E.coli"
 /note="Organ: Developing seed; Vector: pluescript SK-;
 Site_1: EcoRI; Site_2: XhoI"
 BASE COUNT 64 a 56 c 70 g 66 t 4 others
 ORIGIN
 Query Match
 Best Local Similarity 13.3%; Score 252.8; DB 10; Length 260;
 Matches 254; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 OY 411 aagagcgcgaggaaggaaggaagcgagatgcatgtagtatacgaacgtcgt 470
 Db 1 AAGAGCGCGGAGGAAGGAAGGAAGCGGATGCTACGTTTACGTAATCGACCGT 470
 OY 471 tccagctcatcgagagcgagagagtcacatgctcgcgacgaatctcaacaag 530
 Db 61 tccagctcatcgagagcgagagagtcacatgctcgcgacgaatctcaacaag 530
 61 tccagctcatcgagagcgagagagtcacatgctcgcgacgaatctcaacaag 120

ACCESSION LERNA052 LFRA Arabidopsis 663 bp DNA
 VERSION A0958056 sequence.
 KEYWORDS A0958056.1 GI:6785757 GSS 29F-JAN-2000
 SOURCE GSS: linear
 ORGANISM Arabidopsis thaliana LERNA052, DNA
 thale cress.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots;
 Rosidae: eurossids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE Bell,C.R., Lin,X., Pal,G., Bannstead,M., Bowman,C., Ullrichbach,T.,
 (1998) 1 to 6633
 AUTHORS Feldhym,T., Liang,F., Cressy,T.F. and Fraser,C.M.
 TITLE Genomic survey sequencing of Landsberg erecta ecotype of
 Arabidopsis thaliana and identification of sequence-based
 polymorphisms
 JOURNAL Unpublished (2000)
 COMMENT The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: atel@igf.org
 For additional information, see <http://www.tigr.org/tdb/at/at.html>

REFERENCE
Singh, J.A., Mahul, K., Couroux, P., De Moors, A., Harris, L.J., Hattori
Expressed Sequence Tags from Cold-Stressed Maize Seedlings
Contact: Singh, J.A.
Eastern Cereals and Oilseed Research Centre
960 Carling Avenue, Canada
Tel: (613) 759-1662
Fax: (613) 759-1701
Email: singhja@agr.ca

FEATURES
source
location/Qualifiers
1. 662
/organism="Zea mays"
/cultivar="CO328"
/db_xref="taxon:4577"
/clone="Zm04_05g08"
/clone_lib="Zm04.AAFC-ECORC_cold_stressed_maize_seedlings"
/tissue_type="leaf, crown"
/note="Vector: Bluescript SK-/XhoI-ECOR1; Site1: Eco RI;
Site2: Xho I; Lower temperature 50 C / hour from 22 to
, photoperiod 16 hours. Light intensity was 125 ue-1."

[illegible]

TITLE Public Soybean EST Project
JOURNAL unpublished R/Public Soybean EST Project
COMMENT Contact: Shoemaker
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available through: Resgen, Invitrogen Corp. 2130
 South Memorial Parkway Huntleyville, AL 35801 for further information
 call: (800)-533-4363 or contact: coudre@resgen.com web site:
 www.resgen.com
FEATURES Seq Primer: -40RP from GIDCO
 High quality sequence stop: 412.
 Localn/Qualifiers
 1. .372
 /organism="Glycine max"
 source
 -0-01036-7949"

[illegible]

us-09-623-514a-1.rst

Db 181 AGCTACACGCTTCCGCCCTTGTCATTTTCAGTTGACACATGTTGGCTTCCCTATCTT 240
QY 748 atatcagaacgttttgcattcttcttcataatatatcacacagaaaggtttgat 240
Db 241 ATTTATGATGATCTGTGTCACCTGCTTCATATCATTTTACACACTGTAATGTATAT 807
QY 808 ccagtttaagtaacccaaagttgatcttgctttttatcaagtgtaactttaagtc 807
Db 301 CCGTCTGCTGATCTTCTTATGCTGATGTTCTGGGNGTGTCTGGTTTTTGTGATGTTT 300
QY 868 ctcaacttgatgtgtgttgccaagtttgcttcataagccaaactagatagatagcaaa 867
Db 361 ATTGATGATCAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 360

3003

-60-5n

MSU-DOE Plant Research Laboratory
Michigan State University, Plant Biology Bldg. 6
MSU-DOE-PRU, Michigan State University, Plant Biology Bldg. 6
Lansing, MI 48924
Tel.: 517-353-0854
Fax: 517-353-9168
Email: 223tton@dm.c1.msu.edu
Seq primer: T7
Location/Qualifiers

```

/strain="Columbia"
/db_xref="taxon:3702"
/clone="E6B27"
/clone_1b="CD4-13"
+issue type="seedling hypocoty1"
+issue type="seedling hypocoty1"

```

/cys-stage="3 day-old
 /dev-vector: p Bluescript SK-; Site: EcoRI.
 /note="vector: p Bluescript SK-; Site: EcoRI.
 EcoRI; using 5 ug of polyadenylated mRNA from 3 day-old
 Arbidopis chaliana (Columbia), seeding hypocotyls as
 template and oligo (dT) as primer, first strand synthesis
 was catalyzed by Moloney murine leukemia virus reverse
 transcriptase (Pharmacia). Second strand cDNA was made
 using the procedure of Gubler and Hoffman (1983) except
 that DNA ligase was omitted. After the second strand
 reaction, the ends of the cDNA were made blunt with K1
 fragment and EcoRI/NotI adapters (Pharmacia) were ligated
 to each end. The cDNA was purified from unligated
 adapters by spin-column chromatography using Sephadex
 5-300 and size-fractionated on a 5-1 kb) were removed
 mini-gel. Size selected cDNAs (0.5 - 1 kb) were removed
 from the gel using agarase (New England Biolabs),
 phenol:chloroform extracted and precipitated using 0
 NaOH (PH 7)/ethanol. A portion of each cDNA
 size-fraction (0.1 ug) was co-precipitated with 1 ug
 lambda2apii (Stratagene/EcoRI digested, dephosphorylated
 arms and then ligated in vitro using gigaquick II
 ligation mix was packaged in vitro using gigaquick II
 although first strand cDNA synthesis was initiated in
 dt, almost all of the cDNAs begin 8-10 bp from the p
 tail. The reason for the loss of the poly-A tail is
 likely due to lower than anticipated nucleotide before the
 addition of linkers (3' ex instead of 5'-3' pol
 When this library is used please reference the ABCRC
 et al. (1993) Cell 72:427-441."

DB	BASE COUNT	150 a	142 c	0:
	ORIGIN			
	Query Match	2.7%	Score 141.8;	DB 9;
	Best Local Similarity	95.48;	Pred. No. 1.8e-087;	Length 629;
	Matches 146;	Conservative	Mismatches 7;	Indels 0;
				Gaps 0;
OY	3453	ttcaatttgcagcagatataa	ctctgttcaggaaccta	cgacatccttgcgaagcg
		11111111111111111111	11111111111111111111	11111111111111111111
		ttatattatagacacattat	aaattccctttgttcga	acattcctcttggaaagcgc
		11111111111111111111	11111111111111111111	11111111111111111111
		ttatattatagacacattat	aaattccctttgttcga	acattcctcttggaaagcgc
		11111111111111111111	11111111111111111111	11111111111111111111
DB	72	ttatattatagacacattat	aaattccctttgttcga	acattcctcttggaaagcgc
		11111111111111111111	11111111111111111111	11111111111111111111
OY	3513	acctctctatgcattgaaag	atgcttggaagcttcaat	tcacattatgctgac
		11111111111111111111	11111111111111111111	11111111111111111111
		acctctctatgcattgaaag	atgcttggaagcttcaat	tcacattatgctgac
		11111111111111111111	11111111111111111111	11111111111111111111
DB	132	attcttcattatgctcattg	aaagagcttgcgaagctt	caatttcctcaatttat
		11111111111111111111	11111111111111111111	11111111111111111111
		attcttcattatgctcattg	aaagagcttgcgaagctt	caatttcctcaatttat
		11111111111111111111	11111111111111111111	11111111111111111111
OY	3573	tcgcgcattcttaactgctt	ctctcccaaccttggct	3605
		11111111111111111111	11111111111111111111	11111111111111111111
		tcgcgcattcttaactgctt	ctctcccaaccttggct	3605
		11111111111111111111	11111111111111111111	11111111111111111111
DB	192	tctgcacattcttaactgct	ctctcccaaccttggct	224
		11111111111111111111	11111111111111111111	11111111111111111111
		tctgcacattcttaactgct	ctctcccaaccttggct	224
		11111111111111111111	11111111111111111111	11111111111111111111

RESULT 3
AA042298
LOCUS

629 bp mRNA linear EST 19-SEP-1997

us-09-623-514a-1.rst

See 1
on May 13 10:17:46 2002

DEFINITION 24635 CD4-13 Arabidopsis thaliana cDNA clone E6B27, mRNA sequence.
ACCESSION AA042298
VERSION AA042298.1 GI:2414087
KEYWORDS EST
SOURCE thale cress
ORGANISM Arabidopsis thaliana; Streptophyta; Embryophyta; Tracheophyta;
Eukaryota; Viridiplantae; Eudicotyledons; Core eudicots;
Spermatophytes II; Brassicales; Brassicaceae; Arabidopsi

REFERENCE
AUTHORS Newman, T., deBuijn, F.J., Green, P., Keegstra, K., Kende, H., McIntosh
1 (bases 1 to 629)
'E', and Somerville, C.
Genes galore: a summary of methods for accessing results from
large-scale partial sequencing of anonymous Arabidopsis cDNA clones
Plant Physiol. 106, 1241-1255 (1994)

TITLE
JOURNAL On Sep 19, 1997 this sequence version replaced gi:1520456.
MEDLINE Contact: Thomas Newman Laboratory
COMMENT MSU-DOE Plant Research Laboratory
MSU-DOE Plant Research Laboratory
MSU-DOE Plant Research Laboratory
MSU-DOE Plant Research Laboratory
Lansing, MI
Tel: 517-353-0854
Fax: 517-353-9168
Email: 22313tcm@dm.cl.msu.edu
Seq primer: 17
Location/Qualifiers

FEATURES
SOURCE
1. 629
Location/Qualifiers
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="E6B27"
/clone="CD4-13"
/tissue="seedling hypocotyl"
/dev_stage="3 day-old"
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
EcoRI; Using 5 ug of polyadenylated mRNA from 3 day-old
Arabidopsis thaliana (Columbia) seedling hypocotyls as
template and oligo d(T) as primer, first strand synthesis
was catalyzed by Moloney murine leukemia virus reverse
transcriptase (Pharmacia). Second-strand cDNA was made
using the procedure of Gubler and Hoffman (1983) except
that DNA ligase was omitted. After the second strand
reaction, the ends of the cDNA were made blunt with Klenow
fragment and EcoRI/NotI adapters (Pharmacia) were ligated
to each end. The cDNA was purified from unligated
adapters by spun-column chromatography using sephacryl
S-300 and size-fractionated on a 1% low melting point
mini-gel. Size selected cDNAs (0.5 - 1 kb) were removed
from the gel using agarase and precipitated using 0.3 M
phenol:chloroform extracted and precipitated with 1 ug of
NaOAc (pH 7)/ethanol. A portion of each cDNA
size-fraction (0.1 ug) was co-precipitated with 1 ug of
lambdaZapII (Stratagene/EcoRI digested, dephosphorylated
arms and then packaged in vitro using Gigapack II gold
ligation mix was packaged in vitro. We have determined that
although first strand cDNA synthesis was initiated using
at almost all of the cDNAs begin 8-10 bp from the poly-A
tail. The reason for the loss of the poly-A tail is most
likely due to lower than anticipated levels
during the Klenow repair of ragged ends of 5'-3' poly
addition of linkers (3'-5' exo instead of 5'-3' pol)
When this library is used please reference the ABRC and:
Kieber, J. et al. (1993) Cell 72:427-441.

1026b on a part

BASE COUNT
ORIGIN

150 a 142 c 164 g 157 t
17.7% Score 336.4; DB 9; Length 629;
Query Match Best Local Similarity 94.8%; Pred. No. 2.7e-74;
Matches 381; Conservative 0; Mismatches 10; Indels 5; Gaps 3;

OY 1047 TGCATGTATACGGAAGGATGGGCTGCTGTCATATTCGAAACTGATCTATATACCGG 1106
DB 1 TGCATGTATACGGAAGGATGGGCTGCTGTCATATTCGAAACTGATCTATATACCGG 60
OY attcatgagattataatagacacataataatcctatctgtagaactcaagaatc 1166
DB 1107 attcatgagattataatagacacataataatcctatctgtagaactcaagaatc 1226
OY 61 ATTCATGGGATTTATATATGAAACATATATTAATCTATATGACGAACTCAAAATTT 180
DB 1167 ttgaaagcgatctctatcatatgcatatgaaagatgtgaaagcttcaagttc 1286
OY 121 TTGAAAGCGCATCTCTATATGCTATGTAAGAGGTGTAAGCTTCAGTTCCAAATTT 240
DB 1227 atatgtgctctgcatgttctacatgctctcctccacctgtgttaacatgtgcag 1346
OY 181 ATATGCTGCTCTGATGCTTCTACTGCTCTTCTTCCACCTTGTAAACATATTTCCAG 300
DB 1287 gctctctgctgagatgctgcatatctcaaaagattgtggaatgcaaaatgtg 1403
OY 241 GCTTCTGCTCTCGGATGCTGATATTAATTAACAAAGATGGTGAATGCAAAAGTCTGG 360
DB 1347 agatctact-gagaaatgtgataatgctgctcacaatg--atgttgcacata 360
OY 301 AGATTACTGGAGAAATGTGGAATATGCTCTGCTATTAATGGAATGGTCCGACATATTA 1443
DB 1404 C--ttccgigtcttgccagcagacataccaaagacactgcgc 1402
OY 361 CCTTCCCGCTCTTGCACCAAGAAATACCCAAAGACACCC 402

Mon May 13 10:17:45 2002

us-09-623-514a-1.nrl

Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 03:27:01 : Search time 150.64 Seconds
(without alignments)
3104.661 Million cell updates/sec

Title: US-09-623-514A-1
Sequence: 1 attcttagctctctctc.....ttaagtgtaagcgccgc 1904

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 383533 seqs, 122816752 residues
Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA:*
1: /cgn2_6/prodata/2/lna/5A.COMB.seq:*
2: /cgn2_6/prodata/2/lna/5B.COMB.seq:*
3: /cgn2_6/prodata/2/lna/5A.COMB.seq:*
4: /cgn2_6/prodata/2/lna/5B.COMB.seq:*
5: /cgn2_6/prodata/2/lna/PCTUS.COMB.seq:*
6: /cgn2_6/prodata/2/lna/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	336.4	17.7	629	4	US-09-103-754A-3
2	141.8	7.4	1650	4	US-09-103-754A-2
3	141.2	7.4	1976	3	US-09-165-042-2
4	57	3.0	4011	1	US-08-121-057-3
5	57	3.0	4011	2	US-08-509-187D-3
6	57	3.0	4011	2	US-09-121-396-3
7	57	3.0	4011	2	PCT-US93-09704A-3
8	57	3.0	4079	1	US-08-121-057-2
9	57	3.0	4079	2	US-08-509-187D-2
10	57	3.0	4079	2	US-09-121-396-2
11	57	3.0	4079	2	PCT-US93-09704A-2
12	51.6	2.7	7218	1	US-08-232-463-14
13	50	2.6	2040	3	US-09-165-042-4
14	43	2.3	18596	4	US-09-118-448-11
15	35.6	1.9	248	4	US-09-007-005-32
16	35.6	1.9	248	4	US-09-244-796-32
17	35.6	1.9	277	4	US-09-007-005-3
18	35.6	1.9	277	4	US-09-244-796-3
19	35.6	1.9	2367	4	US-09-056-556-201
20	35.6	1.9	4411529	4	US-09-103-840A-1
21	34.8	1.8	906	4	US-08-765-856-3
22	34.8	1.8	906	4	US-08-935-009A-3
23	34.8	1.8	908	4	US-08-765-856-1
24	34.8	1.8	908	4	US-08-935-009A-1
25	34.4	1.8	53526	3	US-08-658-136-2
26	34.4	1.8	53527	3	US-08-658-136-1
27	33.8	1.8	361	4	US-09-018-584A-9

C 28	33.8	1.8	2397	1	US-07-891-942G-11	Sequence 11, Appl
C 29	33.6	1.8	1294	4	US-09-312-038-4	Sequence 4, Appl
C 30	33.6	1.8	2289	4	US-09-312-038-3	Sequence 3, Appl
C 31	33.4	1.8	279	2	US-08-623-906A-3	Sequence 3, Appl
C 32	33.4	1.8	289	4	US-09-007-005-17	Sequence 17, Appl
C 33	33.4	1.8	289	4	US-09-244-796-17	Sequence 14, Appl
C 34	33.4	1.8	7218	1	US-08-232-463-14	Sequence 14, Appl
C 35	33	1.7	80595	4	US-09-078-294-3	Sequence 3, Appl
C 36	32.8	1.7	1298	3	US-08-948-705-3	Sequence 221, App
C 37	32.6	1.7	5785	3	US-08-480-640A-221	Sequence 221, App
C 38	32.6	1.7	5785	4	US-08-688-968C-221	Sequence 221, App
C 39	32.6	1.7	5785	4	US-08-488-237A-221	Sequence 221, App
C 40	32.4	1.7	370	2	US-08-332-766A-8	Sequence 8, Appl
C 41	32.4	1.7	5000	3	US-09-104-070-1	Sequence 1, Appl
C 42	32.4	1.7	80246	4	US-09-078-294-4	Sequence 4, Appl
C 43	32.2	1.7	446	2	US-08-332-766A-26	Sequence 26, Appl
C 44	32.2	1.7	597	2	US-08-332-766A-19	Sequence 19, Appl
C 45	32.2	1.7	1383	2	US-08-630-822A-82	Sequence 82, Appl

ALIGNMENTS

RESULT 1
US-09-103-754A-3
Sequence 3, Application US/09103754A
Patent No. 6344548
GENERAL INFORMATION:
APPLICANT: Farsese, Robert
APPLICANT: Cases, Sylvaine
APPLICANT: Smith, Steven
APPLICANT: Erickson, Sandra
TITLE OF INVENTION: Diacylglycerol O'-acetyltran
TITLE OF INVENTION: sterase
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Bozicevic & Reed
STREET: 285 Hamilton Avenue, Suite 200
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/103,754A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Field, Bret E
REGISTRATION NUMBER: 37,620
REFERENCE/DOCKET NUMBER: 6510-105p
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650 327 3400
TELEFAX: 650 327 3231
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
SEQUENCE: 629 base pairs
LENGTH: 629 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-103-754A-3
Query Match 17.7% Score 336.4; DB 4; Length 629;

100

2-72-042-2

us-09-623-514a-1.rn

```

Query Match          7.4%  Score 141.2; DB 3; Length 1976;
Best Local Similarity 54.2%  Pred. No. 1,2e-33; Indels 3; Gaps 1
Matches 309; Conservative

```

[illegible][illegible]

RESULT 4
US-08-121-057-3
US-08-121-057-3

1 Patent No.:
2 GENERAL INFORMATION: TA-YUAN
3 APPLICANT: CHANG, CATHERINE C. Y.
4 APPLICANT: CHANG, CATHERINE A. CHOLESTEROL
5 TITLE OF INVENTION: ACYL COENZYME A: CHOLESTEROL
6 TITLE OF INVENTION: ACYLTRANSFERASE (ACAT)
7 NUMBER OF SEQUENCES: 4
8 CORRESPONDENCE ADDRESS:
9 ADDRESS: LAHVE & COCKFIELD
10 STREET: 60 STATE STREET, SUITE 510
11 CITY: BOSTON
12 STATE: MA
13 COUNTRY: USA
14 ZIP: 02109
15
16 COMPUTER READABLE FORM:
17 MEDIUM TYPE: floppy disk
18 MEDIUM TYPE: IBM PC compatible
19 COMPUTER: IBM PC-DOS/MS-DOS
20 OPERATING SYSTEM: PC-DOS/MS-DOS
21 SOFTWARE: ASCII text
22
23 CURRENT APPLICATION DATA:
24 APPLICATION NUMBER: US/08/121,057
25
26 FILING DATE:
27 CLASSIFICATION: 800
28 ATTORNEY/AGENT INFORMATION:
29
30

NAME: LAMPORT HAMMITTE, ANN.
REGISTRATION NUMBER: 34,858
REFERENCE/DOCKET NUMBER: DCI-033CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-2700
TELEFAX: (617) 227-5941 3:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 4011 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
FEATURE:
NAME/KEY: CDS
LOCATION: 1397..3046
US-08-121-057-3

Query Match	Similarity	Pred No. 2.00	Indels	Gaps
Best local	46.68	0	Mismatches 285	
Matches 257	Conservative			

[illegible]

RESULT 5
 US-08-509-187D-3 Application US/08509187D
 : Sequence 3, Application US/08509187D
 : Patent No. 5834283
 : GENERAL INFORMATION: Catherine C.Y.
 : APPLICANT: Chang, Ta-yuan and Chang, Catherine C.Y.
 : TITLE OF INVENTION: Cholesterol Acyltransferase
 : NUMBER OF SEQUENCES: 9
 : CORRESPONDENCE ADDRESS:
 :
 :

Mon May 13 10:17:45 2002

ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: 31-JUL-1995
FILING DATE: 31-JUL-1995
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Lamport Hamillite, Ann
REGISTRATION NUMBER: 34,858
TELEPHONE: (617)227-7400
TELEFAX: (617)227-7400
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 4011 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1397..3046
US-08-509-187D-3

Query Match
Best Local Similarity 3.0%; Score 57; DB 2; Length 4011;
Matches 257; Conservative 0; Mismatches 285; Indels 9; Gaps 2;

OY 997 tatttatggtcgtcccatggtggtatcagccagttatcaggtctgcatgata 1056
DB 2291 TACTTCTTATTGCTCTTACCTTATCTACCTGACGATATCCAGGAATCCACGTGA 2350
OY 1057 cggaaagggttggtggtcgtcgaatttgcaaaactggtcatatcccgatcagtgga 1116
DB 2351 AGATGGGATATGCGCTATGAGATTGACAGAGCTTGGCTTCTTCTATGCTGAC 2410
OY 1117 ttataaagaacaataataatcctatgcaaggaactcaagcatccttgaaaggc 1176
DB 2411 TACATCTTGAAGGCTTTGTCCTTGTGCTTGTGCAATATCAACAGAGCCCTTCAGC 2470
OY 1177 gatctctatagcattgaaagggttgaaagcttcagttcccaatttatatgtag 1236
DB 2471 GCTCGTGT---TCGCTCTATGATGATTTTAACTCCATCTTCCAGGCTGCTGATTCG 2527
OY 1237 cctctgcatgctctatgctctcctccaccccttggttaacatattggaagcctctgc 1296
DB 2528 TTCCCTTAACTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTC 1356
OY 1297 ttctggagatcgtgaattcgaagaatggtggaatgcaaaagggtggaagatcagc 1366
DB 2588 TTTGGGTGACAGATGTTCTATTAAGATTGGGACACTCCAGCTCACTCACTCACTATAT 2647
OY 1357 agaatttgaatatgctgctcctcctcctcctcctcctcctcctcctcctcctcctc 2647
DB 2648 ACAACCTGATATGTTGCTCATGACTGCTATATTTACTATGCTTCAAGAGACTTTTC 2707
OY 1411 tgcctgagcagaacatacaagaacacacacacacacacacacacacacacacacacacac 2707
DB 2708 TGGTTTTCCTCAAGATTTCAAAATCTGCTCCATGTTACTGCTTCTGCTGATCTCT 2767

US-09-623-514a-1.inl

OY 1471 gtcttcagtagcctacgacagcagctcctgctcctcctcctcctcctcctcctcctcctc 1530
DB 2768 GTAGTACACGAAATATGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGT 1541
OY 1531 gggatttctgt 1541
DB 2828 CTTCTATGTT 2838

Page 4

RESULT 6
US-09-121-396-3
Sequence 3, Application US/09121396
Patent No. 5968749
GENERAL INFORMATION:
APPLICANT: CHANG, TA-YUAN
TITLE OF INVENTION: ACTYL COENZYME A: CHOLESTEROL
TITLE OF INVENTION: ACTYL COENZYME A: CHOLESTEROL
NUMBER OF SEQUENCES: ACTILTRANSFERASE (ACAT)
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 STATE STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/121,396
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/08/509,187
FILING DATE: 07/31/95
ATTORNEY/AGENT INFORMATION:
NAME: LAMPORT HAMILLITE, ANN
REGISTRATION NUMBER: 34,858
TELEPHONE: (617) 227-2700
TELEFAX: (617) 742-4214
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 4011 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-121-396-3

Query Match
Best Local Similarity 3.0%; Score 57; DB 2; Length 4011;
Matches 257; Conservative 0; Mismatches 285; Indels 9; Gaps 2;

OY 997 tatttatggtcgtcccatggtggtatcagccagttatcaggtctgcatgata 1056
DB 2291 TACTTCTTATTGCTCTTACCTTATCTACCTGACGATATCCAGGAATCCACGTGA 2350
OY 1057 cggaaagggttggtggtcgtcgaatttgcaaaactggtcatatcccgatcagtgga 1116
DB 2351 AGATGGGATATGCGCTATGAGATTGACAGAGCTTGGCTTCTTCTATGCTGAC 2410
OY 1117 ttataaagaacaataataatcctatgcaaggaactcaagcatccttgaaaggc 1176
DB 2411 TACATCTTGAAGGCTTTGTCCTTGTGCTTGTGCAATATCAACAGAGCCCTTCAGC 2470
OY 1177 gatctctatagcattgaaagggttgaaagcttcagttcccaatttatatgtag 1236

Page 6

Y	Conservative	NO. 2, 6e-07;	Indels	Gaps
997	tatttcattgctgcctccacattgtgtatcagccaaagtatccaratt	0; Mismatches 285;	9;	2
2359	TACTTCTTATTTT			

RESULT 9

05-08-509-187D-2 CDNA

Sequence	Matches	Indels	Gaps
997 ttttcatggtcgtccccaattgtattac	285;	9;	3

Db 2359 TACTCTTATTTCCTGACCTTATCTTACCGGTGACAGCTATCCAGGAATCCACAGT 1056
 Oy 1057 CGAAGAGGCTTGGTGGTCGTCGTCATCTTGGCAAACTGGTCAATATCCAGGAATCCACAGT 1056
 Db 2419 AGATGGGGTTATGTTGCTATGAAAGTTTGGACAGGCTTGGTCTCTCTTTTCTATGCTGATC 2418
 Oy 1117 TTATAATAGACAATATATAATACTCTATCTGCGAAGCCCAAGCATCCTTGAAGGC 1116
 Db 2479 TACATCTTTTGAAGGGCTTGTCTGCCCCCTGTTTGGAAATATCAACAGAGCCCTTCAAGC 2478
 Oy 1177 GATCTTCAATATCTATGAAAGAGTGTGTAAGCTTCAAGCTTCCAAATATTAATGATG 1176
 Db 2539 GCTCGGTG---TCTGCTCTATGTTGATTTTAACTCCATCTTCCAGAGTCTGATCTTC 2538
 Oy 1237 CTCTGATGTTCTACTGCTCTCTCCACCTTGGTTAAACAATCTGGAGAGCTCTC 2536
 Db 2596 TTCTCTACTTTTCTTGGCTTTTGGACCTGCTGTCATGCTTCTCTAGATGTTTACGC 2595
 Oy 1297 TTCGGAGTCTGTAATCTCAAGAAAGTGGTGAATGCAAAAGTGTGGAGATTAATC 1296
 Db 2656 TTTGGTACACAGATGTTCTATTAAGSATTGTGGACATCCACGCATACTCCAACTATTAAT 2655
 Oy 1357 AGAATGAGGAATATCCCTGATCAATAAGATGGTGAATGCAAAAGTGTGGAGATTAATC 1356
 Db 2716 AGAACCTGGAAATGTGGTGTCAATGCTGGCTAATTTACTGCTTCAAGGACTTTTTC 2715
 Oy 1411 TGGTTCGAGCAAGATATCAAGACATCGCATATTTACTGCTTCAAGGACTTTTTC 2710
 Db 2776 TGGTTTCTTCACAGAAATATCAAAATCTGCTGCATGTTACTCTCTTCTGATATCTG 2775
 Oy 1471 GTCTTCAATAGTATGATCGAGTCAGTCTCTGTCTGCTCTCTCAAGCTATCTGCT 1470
 Db 2836 GTACTACACGAATATGCTTGGCTGTTGCTTGAAGCTTTTCTATCCGCTCTCTGCTG 2835
 Oy 1531 GAGGATATCTT 1541
 Db 2896 CTCCTCATGTT 2906

Mon May 13 10:17:45 2002

us-09-623-514a-1.rni

Page 7

[illegible]

Db	2716	AGAACCTGGAAATGTGCTGCTCCATCGGCTATATATACATGCTTACAGACATTTCTC	2715
QY	1411	TGCTGCGCGAAGATACCAAGACGACCTGGCATATCATGCTTCTCTAGCTCTGCA	1470
QY	2776	TGGTTTTTCTCCAAAGAGATTCAAATCTGCTGCGCATGTTAGCTCTTGGTGTATCTGT	2835
Db	1471	GGCTCTAGCGCATGATGATGCTCTCTGCTCTCTCAAGCTATGAGCTTCTT	1530
QY	2836	GTTAGTACAGAAATATGCGCTTGGCTGGTTCGTTGCTGAGCTTTCTCTATCCGTCGCTGTTGCG	2899
Db	1531	GGAGATATGCT	1541
QY	2896	CTCTCATGCTT	2906

RESULT 11
PCT-0593-09704A-2
Application PC/TUS9309704A

```

11 RESULT
1 PCT-US93-09704A-2
2 Sequence 2, Application PCT/TUS9309704A
3 GENERAL INFORMATION:
4 APPLICANT:
5 TITLE OF INVENTION: ACYL COENZYME A: CHOLESTEROL
6 ACYLTRANSFERASE (ACAT)
7 TITLE OF INVENTION: 9
8 NUMBER OF SEQUENCES: 9
9 COMPUTER READABLE FORM:
10 MEDIUM type: floppy disk
11 COMPUTER: IBM PC compatible
12 OPERATING SYSTEM: PC-DOS/MS-DOS
13 SOFTWARE: ASCII Text
14 CURRENT APPLICATION DATA:
15 APPLICATION NUMBER: PCT/US93/09704A
16 FILING DATE: October 12, 1993
17 CLASSIFICATION:
18 PRIOR APPLICATION DATA:
19 APPLICATION NUMBER: U.S. SER. NO. 959,950
20 FILING DATE: October 14, 1992
21 APPLICATION NUMBER: U.S. SER. NO. 121,057
22 FILING DATE: September 10, 1993
23 ATTORNEY/AGENT INFORMATION:
24 NAME: LAMPORT HAMMITTE, ANN.
25 REGISTRATION NUMBER: 34,858
26 REFERENCE/DOCKET NUMBER: DCI-033CP
27 TELECOMMUNICATION INFORMATION:
28 TELEPHONE: (617) 227-2700
29 TELEFAX: (617) 227-5941
30 INFORMATION FOR SEQ ID NO: 2:
31 SEQUENCE CHARACTERISTICS:
32 LENGTH: 4079 base pairs
33 TYPE: nucleic acid
34 STRANDEDNESS: single
35 TOPOLOGY: linear
36 MOLECULE TYPE: CDNA
37 PCT-US93-09704A-2

```



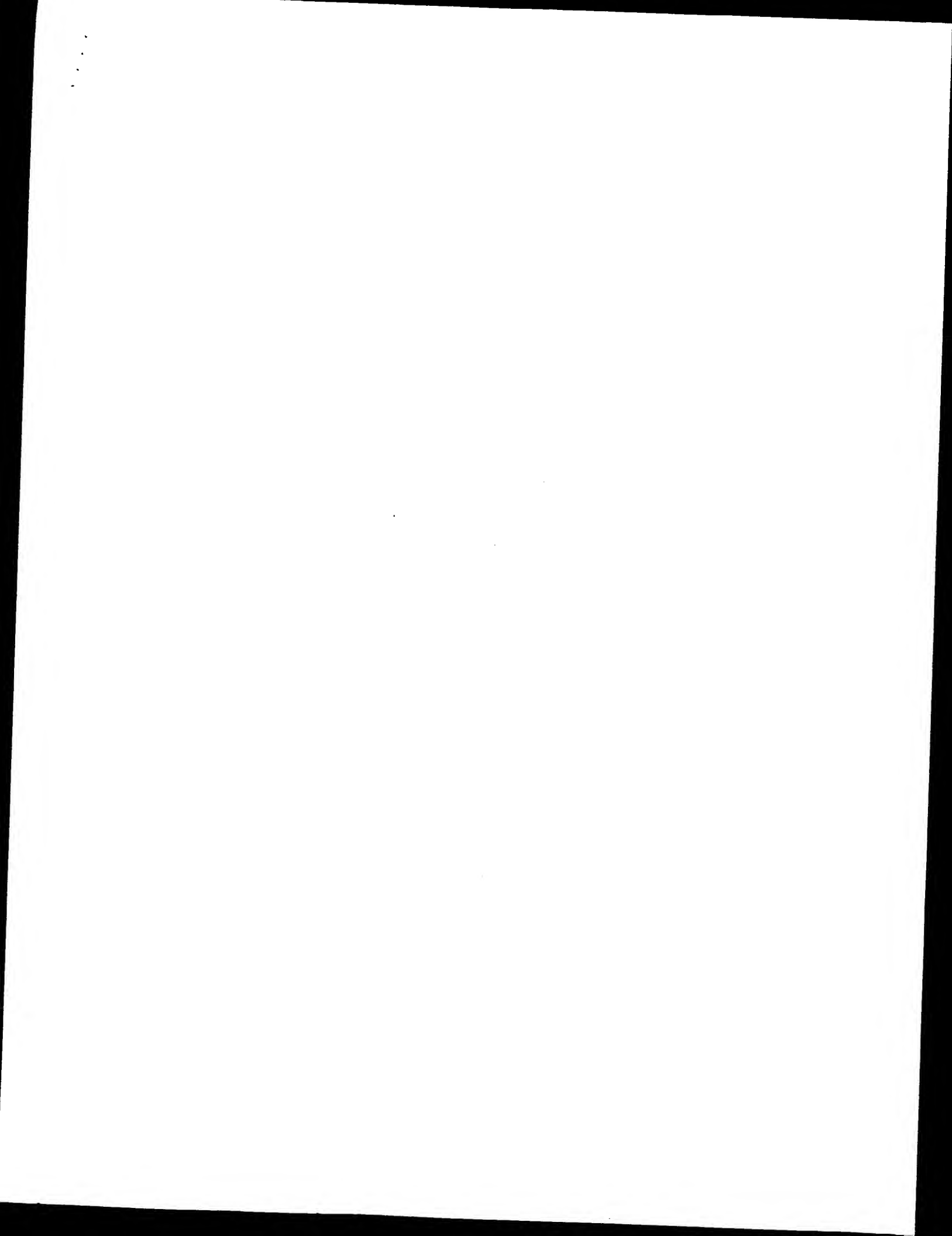
```

QY      632 attcttggtttagttcaagaatcgcgcgaagatgatggccgctttcatggtgtgatataccc 691
Db      248 AAGCTGTGCYAAAGVANGVAGTATYACCGCYCAAGCYCTGTATYTCYCAVANGTATYTTT 1089
QY      692 ttctgatcttcttcttgatcgctgaccttaacggtttgagaattggtactcaagaatacatat 751
Db      188 GATGTGTTATYTCYAAAGCYGCTGTATYTCYACVCAAGCAGTATYTTTAAVCGYCAVAGCY 129
QY      752 cgaagacctgtgcatcttcttcatatattatcacatgacagaggtttgtgatccag 811
Db      128 CTAAGCGATYTCYATYTCYTAATYTCYCAAGVANGVAGTATYTCYCAAGTATYTCYTGATYTCY 69
QY      812 tttagtcaacctaaagtgatgaatctgtttttatcaagtgltcaactttgatgctc 867
Db      68 TATYTCYAAAGCTCAATYTTTGTTAAATYTTGTATYTAATAATYTAAGTATYTAATAATY 13

```

```
Search completed: May 13, 2002, 07:20:45
Job time: 14024 sec
```

Query Match	1.9%	Score 35.6;	DB 4;	Length 248;
Best Local Similarity	22.58;	Pred. No. 0.2;		
Matches	53;	Conservative	77;	Mismatches 106;
				Indels 0;
				Gaps 0.



Mon May 13 10:17:43 2002

us-09-623-514a-1.rge

Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 02:33:51 : Search time 8260.66 Seconds
(without alignments)
4823.359 Million cell updates/sec

Title: US-09-623-514a-1

Perfect score: 1 attcttgccttccttc.....ttatgtgttaagcgccgc 1904

Sequence:

Scoring table: IDENTITY: NUC
Gapop 10.0, Gapext 1.0

Searched: 1797656 segs, 10463268293 residues 3595312

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

GenEmbl:
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_ov:*
21: em_or:*
22: em_ov:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_un:*
28: em_vl:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_inv:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Query Score Match Length DB ID Description

1	1904	100.0	1904	8	ATH238008	ATJ238008 Arabidops
2	1896	99.6	1988	8	ATH131831	AJ131831 Arabidops
3	1895.4	99.5	2005	8	AY054480	AY054480 Arabidops
4	1839.4	96.6	1942	6	AX090345	AX090345 Sequence
5	1839.4	96.6	1942	6	AX090349	AX090349 Sequence
6	1838.6	96.6	1845	6	AF051849	AF051849 Arabidops
7	1146.4	60.2	1537	8	AF251794	AF251794 Arabidops
8	1125.4	59.1	1512	8	AF155224	AF155224 Arabidops
9	954	50.1	1446	8	AF155224	AF155224 Arabidops
10	697	36.6	2099	8	AF129803	AF129803 Nicotiana
11	682	36.3	1964	8	AF129803	AF129803 Nicotiana
12	528	27.7	92721	3	AF221132	AF221132 Caenorhab
13	154.2	8.1	1497	3	CEH19N07	CEH19N07
14	154.2	8.1	11966	3	AF078752	AF078752 Mus muscu
15	145	7.6	1650	10	BC003717	BC003717 Mus muscu
16	145	7.6	1776	10	AY051835	AY051835 Drosophill
17	142.6	7.5	2074	9	AF059202	AF059202 Homo sapi
18	141.2	7.4	1976	9	BC006263	BC006263 Homo sapi
19	139.6	7.3	1493	6	AX090339	AX090339 Sequence
20	139.6	7.3	1895	6	AX090339	AX090339 Sequence
21	139.6	7.3	1998	9	BC015762	BC015762 Homo sapi
22	137.6	7.2	728	10	RN0345014	RN0345014 Rattus no
23	137.6	7.2	1751	10	AF296131	AF296131 Cercopith
24	136.4	7.2	1766	6	AX090340	AX090340 Sequence
25	136.4	7.2	1792	9	AF236018	AF236018 Cercopith
26	133.6	7.0	34948	3	CBRC33823	CBRC33823 Plasmodu
27	129.8	6.8	253305	3	PPMAL387	PPMAL387 Sequence
28	125.8	6.6	275	6	AX090350	AX090350 Sequence
29	122.6	6.4	234	6	AX090350	AX090350 Sequence
30	115	6.0	396	6	AX045795	AX045795 Sequence
31	109.4	5.7	254	6	AX090355	AX090355 Sequence
32	108.2	5.7	267	6	AX090356	AX090356 Sequence
33	103.6	5.4	185932	2	AP003714	AP003714 Oryza sat
34	94.4	5.0	185932	2	AP003714	AP003714 Oryza sat
35	93	4.9	253	6	AX090354	AX090354 Sequence
36	86.4	4.5	40494	2	AC005122	AC005122 Drosophill
37	86.4	4.5	60144	2	AC019798	AC019798 Drosophill
38	86.4	4.5	91685	2	AC005121	AC005121 Drosophill
39	86.4	4.5	170362	3	AC009250	AC009250 Drosophill
40	86.4	4.5	258166	3	AE003652	AE003652 Saccharomyc
41	71.8	3.8	2072	8	SCU51790	U51790 Saccharomyc
42	71.8	3.8	2465	8	SCYNR019W	U51790 Saccharomyc
43	68.6	3.6	2839	8	SCU55383	U55383 Saccharomyc
44	68.4	3.6	325	6	AX090357	AX090357 Sequence
45	68.2	3.6	1673	9	AF053336	AF053336 Chloroceb

ALIGNMENTS

RESULT 1
LOCUS ATH238008 1904 bp mRNA linear JUN-1999
DEFINITION Arabidopsis thaliana mRNA for diacylglycerol acyltransferase.
ACCESSION AJ238008 GI:5123717
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
Zou J., Wei, Y., Jaki, C., Selvaraj, G. and Taylor, D.C.
Direct Submission
Submitted (06-APR-1999) Zou J., Plant Biotechnology Institute,
National Research Council of Canada, 57H OW9, Saskatchewan, CANADA
Location/Qualifiers


```

/function="lipid synthesis"
/codon_start=1
/product="diacylglycerol acyltransferase"
/protein_id="CAB45373.1"
/db_xref="GI:5123718"
/transeq="trna-lys-1"

```

Query Match	100.0%;	Score 1904;	DB 8;	Length 1904
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1904; Conservative	0.000000			

QY	661	gattggccgctttccatgltgtgatatacccttgatcttcttccttgatgccttlaag	720
Db	661	gatttggccgcttttccatggtgtgatattcccttggatcttcttccttgatgccttlaag	720
QY	721	gttcgaagaattggtacttcgaagaatacatcatcagaacctgtgtcaactttttcaatc	780
Db	721	gtttgagaanaattggtacttcgaagaatacattcagaaacctgtgtcaactttttcaatc	780
QY	781	atctacccatctgcagaagttttgatccagtttaagcgaaccccaagggtgatcttgc	840
Db	781	attatcacatgtacagaggtttttgatccagtttaagcgaaccccaagggtgatcttgc	840
QY	841	ttttttcaggtgtcaacttgaatgtgtctcaacttgcattgtgtgctaaagtgtgtct	900
Db	841	ttttttttatcaggtgtcaacttgaatgtgtctcaacttgcattgtgtgctaaagtgtgtct	900
QY	901	tatgtcccaactacagtatagataataagttccctatgcgaatgcagctgtataaggccaatct	960
Db	901	tatgtcccaactacagtatagataataagttccctatgcgaatgcagctgtataaggccaatct	960
QY	961	gaagctctctactacgttaactgttaagaagctgtgcataattatgttgcgtccacattg	1020
Db	961	gaagctctctctactacgttaactgttaagaagctgtgcataattatgttgcgtccacattg	1020
QY	1021	tgttataagccaagtataccaagttctgatatataaggaagaggtgtgggtgcgcgaatg	1080
Db	1021	tgtttatcagccaagattatccacagttctgcattgtatcgaatggaagaggtgtgggtgcgcgaatg	1080
QY	1081	cttcgaaacactgtgtacatactaccggatctcaatgagattataagaacaattataaat	1140
Db	1081	ttttgcaaaaactgtgtacatactaccggatctcaatgagattataagaacaattataaat	1140
QY	1141	cccatgttcaggaactcaagaacatctcttgaaggagatcttcatatgacttgaaga	1200
Db	1141	cccatgtgttcaggaactcaagaacatctcttgaaggagatcttcatatgacttgaaga	1200
QY	1201	gtgttgaagcttctcaatcgaacttaattatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1260
Db	1201	gtgttgaagcttctcaatcgaacttaattatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1260
QY	1261	caacttgtttaaacaatatgtgcagagctctctgtcttgggaacgtggaattctacaa	1320
Db	1261	caacttgttgtttaaacaatatgtgcagagctctctgtcttgggaacgtggaattctacaa	1320
QY	1321	gatttgggtgaatgcanaaagtgltggaagattacttgagaatgttggaatatgtcgttcat	1380
Db	1321	gatttgggtgaatgcanaaagtgltggaagattacttgagaatgttggaatatgtcgttcat	1380
QY	1381	aaatggaatgtgttcgaacatactactccgtgtgtgcagcaagaatacacaagaacac	1440
Db	1381	aaatggaatgtgtgtgacatattatatttcccggtgtgtgcagcaagaatattcacaagaacac	1440
QY	1441	gcaatcatatctgtcttctcgtctcgtcagcttcttcatatgaatcatgactgcgaattcc	1500
Db	1441	gccattatatttgccttctcgtctcgtcagcttcttcatatgaatcatgactgcgaattcc	1500
QY	1501	tgtgtgtctcttcaagctatggagcttctcttgggattatgttccaagtgaccttgtgtctc	1560
Db	1501	tgtgtgtctcttcaagctatggagcttctcttgggattatgttccaagtgaccttgtgtctc	1560
QY	1561	atcacaacaactctacaaggaaggttggctcaaggtgaggaacatgatcttcttgc	1620
Db	1561	atcacaacaactctcacaaggaaggttggctcaaggtgaggaacatgatcttcttgc	1620
QY	1621	atcttcgaagtttgcagaacccgagtgltgtgtcttcttatacagacgtgatgaac	1680
Db	1621	atcttcgaagtttgcagaacccgagtgltgtgtcttcttatacagacgtgatgaac	1680
QY	1681	cgaaagagatcgatgcagaacaacgtttcaaaaatgacttcttcaaacatctatg	1740
Db	1681	cgaaagagatcgatgcagaacaacgtttcaaaaatgacttcttcaaacatctatg	1740

[illegible]

RESULT	3
AY054480	
LOCUS	
DEFINITION	AY054480 2005 bp mRNA linear PLN 05-SEP-2001
ACCESSION	F3P11.5) mRNA, complete cds.
VERSION	AY054480 AY054480.1 GI:15450799
KEYWORDS	FLI_CDNA.
SOURCE	thale cress

ORGANISM	REFERENCE	AUTHORS
<i>Arabidopsis thaliana</i>		
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eustosids II: Brassicales: Brassicaceae: Arabidopsis: 1 (bases 1 to 2005)		
Southwick, A., Karlin-Neumann, G., Nguyen, M., Lam, B., Miranda, M., Palm, C.J., Bowser, L., Jones, T., Banh, J., Carninci, P., Chen, H., Chark, R., Chung, M. K., Hayashizaki, Y., Ishida, J., Kamaya, A., Kawai, J., Kim, C., Lin, J., Liu, S. X., Narasaka, M., Pham, P. K., Sakano, H., Sakurai, T., Satou, M., Seki, M., Shim, P., Shinooka, K., Ecker, J., Theologis, A. and Davis, R. W.		
Submitted (28-Aug-2001)		
Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA		

COMMENT

e-mail for correspondence: arab@sequence.stanford.edu

collection and clustering of RFLP cDNAs (RFLP cDNA) carried out the Arabidopsis Full-length cDNATM (Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamuya, A., Sekurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.).

The Salk, Stanford, PGSC (SSP) Consortium members carried out the sequencing and annotation of the RatL cDNAs: Nguyen, M., Palm, C., Southwick, A., Karlin Neumann, G., Lam, B., Miranda, M., Kim, C., Lin, J., Jones, T., Bani, J., Chen, H., Cheuk, R., Chung, M. K., Yamada, K., Becker, J., Theologis, A. and Davis, R. W.

Seochock, A.V. (SSP/Stanford) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Davis, R.W. (SSP/Stanford) contributed equally to this work as PIs.

FEATURES
SOURCE

Coranismo

111

```
/adb_xret="taxon:3702"
```

chromosome-ITP

11 "TT" - allonge

/cclone="RAFL07-1A-01A

Notes: 1. 2. 3. 4. 5. 6. 7. 8. 9. 10. 11. 12. 13. 14. 15. 16. 17. 18. 19. 20. 21. 22. 23. 24. 25. 26. 27. 28. 29. 30. 31. 32. 33. 34. 35. 36. 37. 38. 39. 40. 41. 42. 43. 44. 45. 46. 47. 48. 49. 50. 51. 52. 53. 54. 55. 56. 57. 58. 59. 60. 61. 62. 63. 64. 65. 66. 67. 68. 69. 70. 71. 72. 73. 74. 75. 76. 77. 78. 79. 80. 81. 82. 83. 84. 85. 86. 87. 88. 89. 90. 91. 92. 93. 94. 95. 96. 97. 98. 99. 100. 101. 102. 103. 104. 105. 106. 107. 108. 109. 110. 111. 112. 113. 114. 115. 116. 117. 118. 119. 120. 121. 122. 123. 124. 125. 126. 127. 128. 129. 130. 131. 132. 133. 134. 135. 136. 137. 138. 139. 140. 141. 142. 143. 144. 145. 146. 147. 148. 149. 150. 151. 152. 153. 154. 155. 156. 157. 158. 159. 160. 161. 162. 163. 164. 165. 166. 167. 168. 169. 170. 171. 172. 173. 174. 175. 176. 177. 178. 179. 180. 181. 182. 183. 184. 185. 186. 187. 188. 189. 190. 191. 192. 193. 194. 195. 196. 197. 198. 199. 200. 201. 202. 203. 204. 205. 206. 207. 208. 209. 210. 211. 212. 213. 214. 215. 216. 217. 218. 219. 220. 221. 222. 223. 224. 225. 226. 227. 228. 229. 230. 231. 232. 233. 234. 235. 236. 237. 238. 239. 240. 241. 242. 243. 244. 245. 246. 247. 248. 249. 250. 251. 252. 253. 254. 255. 256. 257. 258. 259. 260. 261. 262. 263. 264. 265. 266. 267. 268. 269. 270. 271. 272. 273. 274. 275. 276. 277. 278. 279. 280. 281. 282. 283. 284. 285. 286. 287. 288. 289. 290. 291. 292. 293. 294. 295. 296. 297. 298. 299. 300. 301. 302. 303. 304. 305. 306. 307. 308. 309. 310. 311. 312. 313. 314. 315. 316. 317. 318. 319. 320. 321. 322. 323. 324. 325. 326. 327. 328. 329. 330. 331. 332. 333. 334. 335. 336. 337. 338. 339. 340. 341. 342. 343. 344. 345. 346. 347. 348. 349. 350. 351. 352. 353. 354. 355. 356. 357. 358. 359. 360. 361. 362. 363. 364. 365. 366. 367. 368. 369. 370. 371. 372. 373. 374. 375. 376. 377. 378. 379. 380. 381. 382. 383. 384. 385. 386. 387. 388. 389. 390. 391. 392. 393. 394. 395. 396. 397. 398. 399. 400. 401. 402. 403. 404. 405. 406. 407. 408. 409. 410. 411. 412. 413. 414. 415. 416. 417. 418. 419. 420. 421. 422. 423. 424. 425. 426. 427. 428. 429. 430. 431. 432. 433. 434. 435. 436. 437. 438. 439. 440. 441. 442. 443. 444. 445. 446. 447. 448. 449. 450. 451. 452. 453. 454. 455. 456. 457. 458. 459. 460. 461. 462. 463. 464. 465. 466. 467. 468. 469. 470. 471. 472. 473. 474. 475. 476. 477. 478. 479. 480. 481. 482. 483. 484. 485. 486. 487. 488. 489. 490. 491. 492. 493. 494. 495. 496. 497. 498. 499. 500. 501. 502. 503. 504. 505. 506. 507. 508. 509. 510. 511. 512. 513. 514. 515. 516. 517. 518. 519. 520. 521. 522. 523. 524. 525. 526. 527. 528. 529. 530. 531. 532. 533. 534. 535. 536. 537. 538. 539. 540. 541. 542. 543. 544. 545. 546. 547. 548. 549. 550. 551. 552. 553. 554. 555. 556. 557. 558. 559. 560. 561. 562. 563. 564. 565. 566. 567. 568. 569. 570. 571. 572. 573. 574. 575. 576. 577. 578. 579. 580. 581. 582. 583. 584. 585. 586. 587. 588. 589. 590. 591. 592. 593. 594. 595. 596. 597. 598. 599. 600. 601. 602. 603. 604. 605. 606. 607. 608. 609. 610. 611. 612. 613. 614. 615. 616. 617. 618. 619. 620. 621. 622. 623. 624. 625. 626. 627. 628. 629. 630. 631. 632. 633. 634. 635. 636. 637. 638. 639. 640. 641. 642. 643. 644. 645. 646. 647. 648. 649. 650. 651. 652. 653. 654. 655. 656. 657. 658. 659. 660. 661. 662. 663. 664. 665. 666. 667. 668. 669. 670. 671. 672. 673. 674. 675. 676. 677. 678. 679. 680. 681. 682. 683. 684. 685. 686. 687. 688. 689. 690. 691. 692. 693. 694. 695. 696. 697. 698. 699. 700. 701. 702. 703. 704. 705. 706. 707. 708. 709. 710. 711. 712. 713. 714. 715. 716. 717. 718. 719. 720. 721. 722. 723. 724. 725. 726. 727. 728. 729. 730. 731. 732. 733. 734. 735. 736. 737. 738. 739. 740. 741. 742. 743. 744. 745. 746. 747. 748. 749. 750. 751. 752. 753. 754. 755. 756. 757. 758. 759. 760. 761. 762. 763. 764. 765. 766. 767. 768. 769. 770. 771. 772. 773. 774. 775. 776. 777. 778. 779. 780. 781. 782. 783. 784. 785. 786. 787. 788. 789. 790. 791. 792. 793. 794. 795. 796. 797. 798. 799. 800. 801. 802. 803. 804. 805. 806. 807. 808. 809. 810. 811. 812. 813. 814. 815. 816. 817. 818. 819. 820. 821. 822. 823. 824. 825. 826. 827. 828. 829. 830. 831. 832. 833. 834. 835. 836. 837. 838. 839. 840.

THE

T. 2005

1000-1124-3-1001

AT2919450; F3P11.5

162. .1724

1900-1909

Accession number: AF229194.5; F3P11.5

```

/codon start=1

```

T-3700

```
product="dacy[a]vern" 0-2014-1-1
```

protein id="P0AVZ0" name="N-acyltransferase"

T.T/006VWV -PT-UT000-1

```
/u0_Xref="GI:15450800"
```

translation-mutants

MAILDSAGVTTVTENGGEFVDI DBI DDBKDC...

UNNSPSDDVGAPADYRDRIDSVYNDP...
...DDDLKRRKSKSDSSNGILLSG

ATFTVPBPCVDAINDAUGTANLAGDNNGGGDNNGCGCPGCCFCDPAV

..

PLK1DEFWESSRSIRDWPT.FMCTer

ORIGIN

433 d 413 c 447 g 646 t

Query Match	99.58;	Score 1895.4;	DB 8;	Length 2005;
Best Local Similarity	99.98;	Pred. No. 0;		
Matches 1896.				

QY 1 attctcttagctctctcttcaatccgcgtctcttccctctccatctagatcttgctccctt 60
Db 24 ATTTCCTACCTCTCTCTTCCTTCGAATCGCGTCTTTCCCTCTCCATTAATCTCTTCCTTT 83
QY 61 tcaattctctcgcagctctctctctctctctctctcgcagcctctcttcccccagacgtttt 120
Db 84 TCAATTCTCTTCGATGCTCTCTCGATTCTCTCGAGCGCTCTTTCTCCGACGCTTTT 143
QY 121 gctcaaacgcttttcgaaatgacgalttggattctctgtctgtcgttattatagtgagcagg 180
Db 144 CGTCMAAGCCTTTTCGAAATGCGGATTTGGATTCCTGCTGCGCTTACTACGGTGGAGGAG 203
QY 181 aacggttgcgagagatctcgtcagatctctatagcgcttcgttcgaacggaatccagatccgat 240
Db 204 AAGCGTGGCGGAAGTTCGTGATCTGTATAGGCTTCTGTCGACGGAATCGAGATCGGAT 263
QY 241 tctctctaacgagatctctctctctctggttcgcgataaatactctcttcgcgatatgttgg 300
Db 264 TCTTCTAACGAGACTTCTCTCTCTGTTTCGATAATAATCTCTCTCGATGATGTTGGA 323
QY 301 gctccgcgcgacgtaggagatccgattgattccgttgaacgataagacatccagggaca 360

[illegible]

OY	1441	gocattatcatgtcttccctagctcctgcacgtcttcatacgatcagcaatccct	1500
Db	1464	GCCATTATCATGTGCTTTCCTAGTCTCCTGCACATCTTTTCATGATCGATATCAATGGCAGTTTCC	1523
OY	1501	tgtcgtctctccaagcatatggactttcttgaggatgatgtctcagctgaccttgcttc	1560
Db	1524	TGTGCGCTCTTGCAACTATNGGGCTTTCTCTGGGATTAAGTTTAGTGAGGCCCTTTCCTTC	1583
OY	1561	atcacaaactacciaacaagaaggttgctcaacggttgagggaacatgatcttcgcttc	1620
Db	1584	ATCAAACAATACTCTNAGAAGAAAGTTGGCTCAACGGTGGGAACATATCTTCGCTTC	1643
OY	1621	atctctgcattctcggaacaacgcatgtgtgcctctcttatccacgacctgtlgac	1680
Db	1644	ATCTTCTGCATTTTTCGACACACGATGCTGTGCTCTTTATTTACACACCTGTAGAAC	1703
OY	1681	cgaagaagalcagatglatcagtacaacaactcgtltaaaaaatgactcttcaaacatctal	1740
Db	1704	CGAAAAGATCGATGTCATGAAACACTGTTCAAAAAATCACCTTTCTTCAACATCTATAG	1763
OY	1741	gccctcgttgtgacccgcttgatgtgtgtcgtgtgtgtcgtatcgtctaaaagcgaatagtgt	1800
Db	1764	GCCTCGTTGAGATCCTCGTTGATGTTGTGGTGTGCTGTATCTATAACGACAAATAGCTTT	1823
OY	1801	ataccattgaagaagaagaagaattgaagttgttgatcgtacgtaaaaaatlttgttagag	1860
Db	1824	ATAACCATTTGAAGAGAAAGAAATVAGAGTTGTTGATCTGCAGAAAAATTTTGTGTAG	1883
OY	1861	acacgcaaacccgcttgatlttgtlatagtgtataag	1897
Db	1884	ACACGCGAACCCGTTTGATTTTGTATGTGTTAAG	1920
RESULT	4		
AX090345		1942 bp	DNA linear PAT 21-MAR-2001
LOCUS	Sequence 38 from Patent WO0116308.		
DEFINITION	AX090345		
ACCESSION	AX090345.1 GI:13444207		
VERSION			
KEYWORDS	thalae cress.		
SOURCE ORGANISM	Arabidopsis thaliana		
REFERENCE	Eukaryotes: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
AUTHORS	Eukariota: Magnoliophyta; eudicotyledons: core eudicots;		
TITLE	Spermatophyta: Rosidae II; Brassicales; Brassicaceae; Arabidopsids.		
JOURNAL	Rosidae: eurosids I to 1942)		
	1 (bases 1 to 1942)		
	Lassner,M. and van Bennaam,A.		
	Plant sterol acyltransferases		
	Patent: WO 0116308-A 38 08-MAR-2001;		
	MONSANTO COMPANY (US)		
FEATURES	Location/Qualifiers		
source	1..1942		
	/organism="Arabidopsis thaliana"		
	/db_xref="taxon:3702"		
BASE COUNT	461 a 421 c 425 g 635 t		
ORIGIN			
Query Match	96.6%; Score 1839.4; DB 6; Length 1942;		
Best Local Similarity	99.9%; Pred. No. 0;		
Matches 1840; Conservative	0; Mismatches 1; Indels 0; Gaps 0;		
OY	1	attctttagctcttcctccaatcgcgtcttcccttcacataagttctgttccctt	60
Db	99	ATTCTTTAGCTTTCCTTCCTTCAATCCGCTTTCCCTTCATTAAGTTCGTTCCTCT	158
OY	61	tcaattctctcgtacgtctctcgaatctctcgaaccccttcttcgcgaagctgtt	120
Db	159	TCAATTCTCTTCGCAATGCTTTCGATTTCTCTGAACGCTCTTTCCTCCGACGCTGTT	218
OY	121	cgctcaacgcttctcgaatagcgatcttgatctcgtcgtctacagtgtagag	180

D	219	CGTCAACGCTTTTGGAAATGGCATTTTGGATTTCGTGGGCTTACTACGTCACGGAG	278
Q	181	aacggtgcgagagagtttcgtcagatctgtgaagcttcgtcgagcggaatcagatcgcg	240
D	279	AACGGTGCAGGAGAGTTCGTGATCTTGATAGGCTTCGCGACGAATCCAGATTCGGAT	338
Q	241	tcttcaacgactctctctcgtgttcgcgataaattctcctcgcgagtagtggg	300
D	339	TCTCTAACGGACTCTTCTCTCTGTGGTCCGATAATATTCCTCTCGATGATGTTGG	398
Q	301	gctccgcgagctttagggatcogagttatccgtgtgtgaagctgagctcaggaaca	360
D	399	GCTCCGCCGACCTTGGATCGGATTCGATTTCCTGTTAACGATACGCTCAGGGAACA	458
Q	361	gccaatctggcgcgagatataaagctgtgtgcgataataagctgtgtgaagagcggc	420
D	459	GCCAAATTTGGCGGAGATATTAACGCTGTGGCATATATACGTTGGTGGAAAGACGAGC	518
Q	421	ggaagaaggaagggaacgcgcgaatgctgaagttacgatacgcgcgtcttcacgtat	480
D	519	GGAGAAGGAAGGAAGAACCCGATGTCACGTTTACGATTCGACCGTGGTTCACCTAT	578
Q	481	cgagagcgagagagagatcgaactagctccgaagcaatctcaaacagagccatgcgga	540
D	579	CGAGGCGGAGAGAGATCCACTTACGTCGACGCAATTTTAAACAAGCAATCCGCGA	638
Q	541	ttactcaacctctgttagtagttcttattgcgtgaacagtagactcatcgaagaat	600
D	639	TTATTCACCTCTGTAGTAGTTCTTATTGCTGTAACAGATACATCATCATCGAATAAT	698
Q	601	ctatgaagatggtgtgtgatacgaacggaattctgtgttagtcaagatgcgcga	660
D	699	CTATGTAAGTATGTTGCTGTGATCAGAAAGGATTTCTGGTTAGTTCAATGCGTGGCA	758
Q	661	gattggcgcttctcaatgtgtgtatatacccttcgaatcttcccttgctgcgttaag	720
D	759	GATTGGCGCTTTCAATGTTGTGTATTCCTTCGATCTTTCCTTGGCTGCTTTACG	818
Q	721	gttgagaattgtaactcaagaataatcagaacctgtgtcaactcttctcatatt	780
D	819	GTTGACAATTTGTCATTTCAGAAATACATATACAGAACCTTGTCATCTTCTCATATT	878
Q	781	attatcacatpagaagagtttgcatalcgaattgaacccatgaagtagatgtcgt	840
D	879	ATTATCCACATACAGAGGCTTTGTATCCAGTTTACGTACCCCTTAAGGTGATTTGCT	938
Q	841	ttttataagctgtcaacttgatgcctcccaacttgatgtgtgtcgaagtgttct	900
D	939	TTTTATAGGTCACCTTGATGCTTCACCTGATGCTGATGCTGCAAAATTTGGTTCT	998
Q	901	tatgcatacgaactatgacataagaatccctagccaatgacgtatgaagccaatcc	960
D	999	TATGCTCATCTAGTACATATGATACATACATCCCAAGCAATGACGTGATAGGCAATCT	1058
Q	961	gaatgccactcaagtttagcttgaagagctgtgatatctcaatgctgcgtccacatgt	1020
D	1059	GAATGCTCCTACACTGATGCTTGTGAAGAGCTTTGGCATTTTCATGCTGCCCATTTG	1118
Q	1021	tgatcacgccaagttatccagcttcgcagatatacgaagagctgtgtgtgcgtcaa	1080
D	1119	TGTTATCACCAAGTTATCCAGCTTTCGATGATATACGGAAGGTTGGGTGCTGCTCA	1178
Q	1081	tttgcaaacctgtatattcaaccgagatcaatggaattatataagaacaatataat	1140
D	1179	TTTGGCAAACTGCTATATATCACGCGATTCATGCGATTATATATGAAACATATATTAAT	1238
Q	1141	ccattctgcaggaactcaaacgacatcccttgaagagagactctatatgcatatgaaga	1200
D	1239	CTATTGTGCAGAACTCAAAAGCATCTTTGAAGAGGAGATCTTATATGATATGAAGA	1298
Q	1201	gttgtgaagcttcaagtccaattatatgtgtgtgtgcgtcaagttctactgtcttc	1260
D	1299	GTGTGAAGCTTTCAAGTTATATGTTGTGCTGCTCATGTTCTATGCTTCTTC	1358

QY	1261	caacttgggttaacacataltgycagagctctctgcttcggggatcgtaattctaca	132
Db	1359	CACCTTTGGTAAACATATTGGCAGAGCTTCTCTCTCGGCGATCGTAATCTCCAAA	141
QY	1321	gattgltggaatgcanaaagltggtgggaattactgggaatgtaagtatgacctgtcat	1381
Db	1419	GATTGGTGAATGCAAAAGGTGGGAATTACTGGGAATATGGGAATATGGCTGTCAAT	1478
QY	1381	aaatggatgtgtccagataataactctccgltgcttgcagacaagataccaaagacac	1444
Db	1479	AAATGATGCTTGCACATATATACTCCCGCTTGGCGAGCAAGATACCAAGACACTC	1538
QY	1441	ggcatatcatgtgcttctctctagctctgcagctctctcaatgagatgactgcagttcc	1500
Db	1539	GGCATATCATATGCTTCTCTCAATGCTCGAGCTTTTCATGAGACTATGCATGCAATTCCT	1598
QY	1501	tgtctctcttcaagatctagtgagcttctcttgggaattatgttccagtgaccttgcttc	1560
Db	1599	TGCTCTCTTCAAGCTATGAGGCTTCTTGGGATATATCTTCAGAGCTCTTGGCTTC	1658
QY	1561	atccaaactctctcaacaggaagaagtttgctctaacagtggtgggaacatgactctgttc	1620
Db	1659	ATCACAAACTATCTACACGGAAGAGTTTGGCTCAACGGTGGGAACATGATCTTGCTTC	1718
QY	1621	atctctgcatttctggagacaacccgatalgtgtgtctcttcttaccagaccctgataaac	1680
Db	1719	ATCTTCTGCATTTTTCGCAACCCGATGTGTGTCTTTATTTACACGACCTGATGAAC	1778
QY	1681	cgaagaagatcgatgtcatgaacaacagcttctcaaaaatgacttcttcaaatctatg	1740
Db	1779	CGAAAGGATGCATGCTCATGAAACACAGTCTCAAAAATGACTTTCTTCAACATCTATG	1838
QY	1741	gacctgtgtgactccgtctgttgtgtgtgtgtctgtatgtctaaacgacaatagtggt	1800
Db	1839	GGCTGTTGATGCTCCGTTGATGTGTGTGTGTGTGTGATGCTTAAACGACCAATATGT	1898
QY	1801	ataccattgaagaagaagaagaatagattagattgtgtatc	1841
Db	1899	ATAACCATTTGAAGAAAGAACAAATTAGTGTGTATC	1939
RESULT 5			
LOCUS	AX090349	1942 bp	DNA
DEFINITION	Sequence 42 from Patent WO0116308.		linear
ACCESSION	AX090349		
VERSION	AX090349.1	GI:13444210	
KEYWORDS			
SOURCE			
ORGANISM	thale cress.		
	Arabidopsis thaliana		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta		
REFERENCE	Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
AUTHORS	1 (bases 1 to 1942)		
TITLE	Lassner, M. and Van Eenennaam, A.		
JOURNAL	Plant sterol acyltransferases		
	Patent: WO 0116308-A 42 08-MAR-2001;		
FEATURES	MONSANTO COMPANY (US)		
source	location/Qualifiers		
	1..1942		
BASE COUNT	/organism="Arabidopsis thaliana"		
ORIGIN	461 a 421 c 425 g 635 t		
Query Match			
	96.6%;	Score 1839.4;	DB 6;
	Best Local Similarity 99.9%;	Pred. No. 0;	Length 1942;
	Matches 1840;	Conservative 0;	Mismatches 1;
			Indels 0;
			Gaps 0;
QY	1	attctttagcttctcttcaatcgactcttctctcctcatagattcgttctctt	60

JOURNAL Submitted (28-FEB-1998) Institut de Biologie Moléculaire des
Plantes, Centre National de la Recherche Scientifique, 28 rue
Goethe, Strasbourg 67083, France
COMMENT On Jan 24, 2000 this sequence version replaced gi:650202.
FEATURES
Location/Qualifiers
1. 1845
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/tissue_type="hypocotyl"
/clone_lib="ABRC library, catalog number CD4-15"
/dev_stage="3 day-old seedling"
/note="screened with EST, GenBank Accession Number
AA042298"
1. 1845
/gene="DAGAT"
88. 1650
/gene="DAGAT"
/function="catalyzes the acylation of diacylglycerol by
oleoyl-CoA to give triacylglycerol"
/note="acyl-CoA cholesterol acyltransferase homolog"
/product="diacylglycerol acyltransferase"
/protein_id="BAF19262.1"
/db_xref="GI:662553"
/translation="MAIDSGAVTVTENGGEFVDLRLRRKSRSDSSNGLLSSG
DNNSPSDVGADPADVDRIIDSVNDADACTAMLAGDNGGSDNGSGGEGGND
ATFTYRSPVAHRRRRESPLSDAIPKOSHAEFLNLCVAVLIIENLKKYAM
LIRTEDEFSRSLRDLPMCCISIFLAFVTEKLVSEPVLEIHLITIM
TEVILYEVYVTLRCDSAFLSCVTLMELCTIWKLVSYAHSTYRSLANADANPEY
SYVSLKSLAEVMPATLCTOPSYRSCIRGWARFALVLETPMFIROYIN
PIYRNSKHPKGLDLALIBRLKLSVPMVLMCFCEFLHMTNIAELLCGDBE
YKDMNAKSVGVYWRMMMPVHKMWVRHLYPCRSKIPKLIATIIAFLVSAVHELC
IAYCPLEPRLTFLGIMFQVPLVFTITNLOERFSTVGMIMFETICFQPMCVILY
YHDMNRKSKMS"

gene
1. 1845
/gene="DAGAT"
88. 1650
/gene="DAGAT"
/function="catalyzes the acylation of diacylglycerol by
oleoyl-CoA to give triacylglycerol"
/note="acyl-CoA cholesterol acyltransferase homolog"
/product="diacylglycerol acyltransferase"
/protein_id="BAF19262.1"
/db_xref="GI:662553"
/translation="MAIDSGAVTVTENGGEFVDLRLRRKSRSDSSNGLLSSG
DNNSPSDVGADPADVDRIIDSVNDADACTAMLAGDNGGSDNGSGGEGGND
ATFTYRSPVAHRRRRESPLSDAIPKOSHAEFLNLCVAVLIIENLKKYAM
LIRTEDEFSRSLRDLPMCCISIFLAFVTEKLVSEPVLEIHLITIM
TEVILYEVYVTLRCDSAFLSCVTLMELCTIWKLVSYAHSTYRSLANADANPEY
SYVSLKSLAEVMPATLCTOPSYRSCIRGWARFALVLETPMFIROYIN
PIYRNSKHPKGLDLALIBRLKLSVPMVLMCFCEFLHMTNIAELLCGDBE
YKDMNAKSVGVYWRMMMPVHKMWVRHLYPCRSKIPKLIATIIAFLVSAVHELC
IAYCPLEPRLTFLGIMFQVPLVFTITNLOERFSTVGMIMFETICFQPMCVILY
YHDMNRKSKMS"

BASE COUNT 453 a 378 c 421 g 593 t
ORIGIN

Query Match 96.6%; Score 1838.6; DB 8; Length 1845;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1841; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 52 ttccctctcaattcttctgcatgctctcgaatctctctcgaagctcttctcccg 111
DB 1 TTTCCCTCTTCAATTTCTTCTGCAATGCTTCTGATCTCTGATCGCTCTTTCTCCG 60
QY 112 acgctgcttctgcaagctcttcaaatgcaatttgatctctgctgagcttactag 171
DB 61 ACGGCTTTCGTCAAACGCTTTTGAATGCGATTTTGATTTCTGCTGCGTACTAG 120
QY 172 gtgaag 231
DB 121 GTGACGAGGAG 231
QY 222 agatcgatctctcttaagagctctctctctgctgctgcaataatctctcgaat 291
DB 181 AGATCGGATCTTCTTAACGAGCTTCTCTCTGCTTCGATTAATAATTTCTCTCGAT 240
QY 292 gatcttgagagctcccgacgcttggatcgatgattcgcgtgtaagagagagct 351
DB 241 GATGTTGAGAGCTCCCGCCAGCTGATGAGATGATTCCTGTTTAACGATGAGAGCT 300
QY 352 caag 411
DB 301 CAGGAG 411
QY 412 aag 471
DB 361 AAG 471
QY 472 ccagctcag 531
DB 420 ccagctcag 531

DB 421 CCAAGCTCATCGAG 480
QY 532 catcgag 591
DB 481 CATGCGGAGATATTCACGATGCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCT 540
QY 592 atcgaaatctatgaatgagatgagatgagatgagatgagatgagatgagatgag 651
DB 541 ATCGAAATCTTATGAGATATGAGATGAGATGAGATGAGATGAGATGAGATGAG 600
QY 652 tgcgag 711
DB 601 TCGCTGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGAT 660
QY 712 gcttcaagctgag 771
DB 661 GCTTTACGAGTGTGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGAT 720
QY 772 ctcatatctatctacacagagagagagagagagagagagagagagagagagagag 831
DB 721 CTTCATATTTATATACAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGAT 780
QY 832 gatctgcttcttcaag 891
DB 781 GATTCGCTTTTATACAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGAT 840
QY 892 ttgcttctatctacacagagagagagagagagagagagagagagagagagagag 951
DB 841 TTGGTTTCTTATACAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGAT 900
QY 952 gcaatctcgaagctcctactacagagagagagagagagagagagagagagagag 1011
DB 901 GCCAATCTGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGAT 960
QY 1012 cccaatctgag 1071
DB 961 CCCAATCTGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGAT 1020
QY 1072 gctgctaatctcaaatctgagagagagagagagagagagagagagagagagag 1131
DB 1021 GCTGCTAATTTATACAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGAT 1180
QY 1132 tatataatctatctacacagagagagagagagagagagagagagagagagagag 1191
DB 1081 TATATAATCTTATACAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGAT 1140
QY 1192 atgaaag 1251
DB 1141 ATGAAAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGAT 1200
QY 1252 tgccttctcaaccttggatgaacatctgagagagagagagagagagagagagag 1311
DB 1201 TGCCTTCTTCAACCTTGGTGAACATTTGGCAGAGAGAGAGAGAGAGAGAGAGAG 1260
QY 1312 ttctacaag 1371
DB 1261 TTCTTCAAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGAT 1330
QY 1372 cctgttcaatgaag 1431
DB 1321 CCTGTTCAATGAAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGAT 1380
QY 1432 aagaaactcgaatctatctctctctctctctctctctctctctctctctctct 1491
DB 1381 AAAAACAATCTTATACAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGAT 1440
QY 1492 ggaattct 1551
DB 1441 GCAATCTTCTTCT 1500
QY 1552 ttgcttctcaatcaatctatctacagagagagagagagagagagagagagagagag 1560
DB 1501 TTGCTTCTTCAATCTATCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1560

[illegible]

QY	121	ttcagttccaaattatataiytgcgtctcgcatglttcacgtctctccaacttgagt	1271
Db	1038	tttcagttccaaattattacgtctgacctgcgatgtttctactgctcttccacccttgagg	1097
QY	1272	aaaacatatitggcacagctctctcgtcttcgggatctcgtgaacttacaaagaattggagaa	1331
Db	1098	AAACATATTGGCAGACGCTCCTCGCTTGCGGAATCGTAATTCMACAAAGATTGGTGGAA	1157
QY	1332	tgc aaa agctggtaggagatctaactggaagaigtggaatataibcgtgtcataaatagtgagt	1391
Db	1158	TGC AAA AGCGTTGGGAGATTATTGGAGAAATTGTGAATATGCCGTTCAATAATGGATGCT	1217
QY	1392	tggacataatactactcccgtctgttcggcaagaagataaccaagaacactgcgcatalcoat	1451
Db	1218	TCGACATGTAATCTTCCGTGCGCTTCGCAGAAATATACCGAAGAAGTACCGCATTTATTCCT	1277
QY	1452	tgcttccagctcgtcgcagctcttaagtgcatagtacalcgcgaagttccttgttctctt	1511
Db	1278	TGCTTCTTAGTCCTCGCAGCTCTTCAATGAGTATGATCATCGCAGTTCCTTGTCTCTCT	1337
QY	1512	caagctatgggcttcttcttggaaltatgtltaacgtgcttgtgtcttcatacaaatla	1571
Db	1338	CAAACTATGAGCGCTTCTTGSGGATTAATGTTTACGTTGCCCTTTGGATTTATCAACAACYA	1397
QY	1572	tctacaggaagaagttgtgtgcacaagctggggaacaatgactcttcgtctcaactcgtcag	1631
Db	1398	CCTACAAAGAAAGTTTGCGTCCATGAGTGGGAAACATGATATTTCGTTTTACCTTCGCAT	1457
QY	1632	tttcggacaacagcatgtgtgtgcttcttatacacaagactgaataaacccgaagaagatc	1691
Db	1458	TTTGCGACAAACCATGTGTGTGCTTCTTATTATCACGACTTATATACCGAAGGAAGAA	1517
QY	1692	gatgtcat 1699 	
Db	1518	GATGT CAT 1525	
RESULT	8		
LOCUS	AF164434	1512 bp	mRNA linear PLN 30-NOV-1999
DEFINITION	Brassica napus putative diacylglycerol acyltransferase (DGAT1)		
ACCESSION	AF164434		
VERSION	AF164434.1	GI:5579407	
KEYWORDS			
SOURCE			
ORGANISM	Rape.		
	Brassica napus		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
	Rosidae; eurosoids II; Brassicales; Brassicaceae; Brassica.		
REFERENCE	Nykiforuk,C.L., Laroche,A. and Weselake,R.J.		
AUTHORS	Isolation and Characterization of a cDNA Encoding a Second Putative		
TITLE	Diacylglycerol Acyltransferase from a Microspore-derived Cell		
JOURNAL	Suspension Culture of Brassica napus L. cv Jet Neuf (Accession No.		
REFERENCE	Plant Physiol. 121 (3), 1053 (1999)		
AUTHORS	Nykiforuk,C.L., Laroche,A. and Weselake,R.J.		
TITLE	Direct Submission		
JOURNAL	Submitted (30-JUN-1999) Chemistry and Biochemistry, University of		
	Lethbridge, 4401 University Drive, Lethbridge, Alberta T1K 3M4,		
FEATURES			
source	Location/Qualifiers		
	1..1512		
	/organism="Brassica napus"		
	/strain="Cultivar Jet Neuf"		
	/db_xref="taxon:3708"		
gene	/note="microspore-derived cell suspension culture"		
	1..1512		
CDS	/gene="DGAT1"		
	1..1512		

[illegible]

OY 919 gacataagatccctagccaatgcaagcgaatccctgaagtcctactactcgt 978
 ||||||| ||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
 Db 730 GACATTAAGAACCTGGCCATTCACCTGATTAAGGCGATCGTGAATCTCTATAGT 789
 OY 979 agcttgaagcttgcgaatcttcaatgctgcctcccaatgcttgaacgaactat 1038
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
 Db 790 AGCTTGAAGGCTTGGCGATTCATAGCTGCTCCACACTGCTTTATGACGCAAGCTAT 849
 OY 1039 ccacgttcgaatgataaggaaggttggtgctgcgtcaatttgcgaactgata 1098
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
 Db 850 CCAGCTCTCCATGATTCGGAAGGCTTGGGCGCTCGTCAACTGCAAACTGCTCAT 909
 OY 1099 ttccagatcatgaggaattataatagaacataataatccatctgtaagaactca 1158
 ||||||| ||| ||||||| ||||||| ||||||| ||||||| |||
 Db 910 TTCACTGACATCAGGATTTATATAGACCAATATATTAATCTATTGTTAGCAACCA 969
 OY 1159 aagcaatcttgaagagcgatctctataatgctatgaaagagcttgaagcttca 1218
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
 Db 970 AACCATCTCTGCAAGGAGGACCTTCTATGCTATTCGAAGAGTGTGAAGCTTTCAT 1029
 OY 1219 ccaattatagtgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1278
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
 Db 1030 CCAATCTATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1089
 OY 1279 ttggagagcttgcctgctgctgctgctgctgctgctgctgctgctgctgct 1338
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
 Db 1090 TTGGCAGAGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1149
 OY 1339 agtgaggagattactgaggaatggaatggaatggaatggaatggaatggaat 1398
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
 Db 1150 AGCTTGGAGATTTATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 1209
 OY 1399 atatactcccgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1458
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
 Db 1210 GTATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1269
 OY 1459 ctgactctgagctcttcatgagcagcagcagcagcagcagcagcagcagcagc 1518
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
 Db 1270 TTAAGTCCTGCAAGCTTTCATGATTTGATGATGATGATGATGATGATGATGAT 1329
 OY 1519 tgggcttcttggagattatgcttcaagtgcttgccttgccttgccttgccttgcct 1578
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
 Db 1330 TGGGCTTTCATGGAATTTATGATGATGATGATGATGATGATGATGATGATGAT 1389
 OY 1579 gaaagttggttcaagcagcagcagcagcagcagcagcagcagcagcagcagc 1638
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
 Db 1390 GAAAGGTTTGGCTCCATGCTGGAACATGATCTTGGTTCAGCTTCTTCATTTT 1449
 OY 1639 caacgagctgtgcttcttcttcttcttcttcttcttcttcttcttcttct 1698
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
 Db 1450 CACCGATGCTGCTGCTTCTTATTCATGATGATGATGATGATGATGATGATGAT 1509
 OY 1699 tga 1701
 |||
 Db 1510 TGA 1512
 |||

RESULT 9
 AF155224 1446 bp mRNA linear PLN 25-AUG-1999
 LOCUS AF155224
 DEFINITION Brassica napus putative diacylglycerol acyltransferase (DGAT2)
 VERSION AF155224.1 GI:5225381
 KEYWORDS
 SOURCE Brassica napus
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
 Rosidae; eurosids II (bases 1 to 1446)
 REFERENCE
 AUTHORS Nykliforuk, C.L., Laroché, A. and Weselake, R.J.

TITLE Isolation and Sequence Analysis of a Novel cDNA Encoding a Putative
 Diacylglycerol Acyltransferase from a Microspore-derived Cell
 Suspension Culture of Brassica napus L. cv Jet Neuf (Accession No.
 AF155224). (PGR99-123)
 JOURNAL Plant Physiol. 120 (4), 1207 (1999)
 REFERENCE 2 (bases 1 to 1446)
 AUTHORS Nykliforuk, C.L., Laroché, A. and Weselake, R.J.
 TITLE Direct Submission
 JOURNAL Submitted (31-MAY-1999) Chemistry and Biochemistry, University of
 Lethbridge, 4401 University Drive, Lethbridge, Alberta T1K 3M4,
 Canada

FEATURES
 source Location/Qualifiers
 1..1446
 /organism="Brassica napus"
 /cultivar="Jet Neuf"
 /db_xref="taxon:3708"
 /note="microspore-derived cell suspension culture"
 gene
 1..1446
 /gene="DGAT2"
 82..1107
 /gene="DGAT2"
 /c-number="2,3,1,20"
 /product="putative diacylglycerol acyltransferase"
 /codon_start=1
 /protein_id="FAD40881.1"
 /db_xref="GI:5225382"
 /translation="MGISLSTPLAFTVEKLVKQKISEPVVILHVTITVEVLY
 /PVTYLRCDNAFLSGDTMLITCIWKLVSVAHNNYDIRLANSNDKANEYSYS
 LKSLAYEMLAFTLYOFSIPRSRPIKGVAVQAFKLYIFGLMGFIIEOYINIVAN
 SKHPLKGLDLYALEIRVLYKSPVNLVYVLCMPGFHMLNLTALIELCEGDEPKDW
 NAKSVGDYRMNNMPVHRKMWVRHVYFPCLRIRKIPVPAIIIAFVSAVFLHLCYAVC
 RLFINMAFMGIMFQVPLVFTINFLQEFPSVMGNMIFGASGIFQPMCHLYYHDIM
 NRKGSMS"

BASE COUNT 398 a 291 c 278 g 479 t
 ORIGIN

Query Match 50.1%; Score 954; DB 8; Length 1446;
 Best Local Similarity 91.8%; Pred. No. 3e-234;
 Matches 1019; Conservative 0; Mismatches 90; Indels 1; Gaps 1;

OY 594 cgaaaatcttgaagatgctgctgctgctgctgctgctgctgctgctgctgctgct 653
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
 Db 1 CGAAAATCTCATGAGAGTACGTTGCTGATCAGAACTGATGATGATGATGATGATGAT 60
 OY 654 gctgcgagatggcgcgcttctcaatgctgctgctgctgctgctgctgctgctgct 713
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
 Db 61 GCTGCGAGAT- GCCGCTTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 119
 OY 714 cttaagcttgaagaaatgctgctgctgctgctgctgctgctgctgctgctgctgct 179
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
 Db 120 CTTTACCCTGCGAGAAATAGTACTTCAGAAATGATATGATGATGATGATGATGATGAT 239
 OY 774 tcaatattatcacagacagagcttctgtaacagcttcaacagcttcaacagcttcaac 833
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
 Db 180 TCATGCTATATATACCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 299
 OY 834 ttctgcttcttcaagctgctgctgctgctgctgctgctgctgctgctgctgct 893
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
 Db 240 TTCTGCTTCTTATCAGAGTACACGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 953
 OY 894 gtttcttcaatctatctatgctgctgctgctgctgctgctgctgctgctgctgct 953
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
 Db 300 GCTTCTTACCTGATATCTATGATGATGATGATGATGATGATGATGATGATGATGAT 1013
 OY 954 caatctgaagctctccttcaatgctgctgctgctgctgctgctgctgctgctgct 419
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
 Db 360 CAACTCTGAAGCTCTCTACTATGTTAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCT 479
 OY 1014 caatgctgcttcaagcgaagcttcaagcttcaagcttcaagcttcaagcttcaag 1073
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
 Db 420 CAACTGCTTATCAGCAGCTATCAGCTTCCATGATGATGATGATGATGATGATGATGAT 479

OY 1074 tgcgaattgcaaaactgcatatcaccgagatcattggaatttaataagaacata 1133
 DB 480 TCGTCATATTCGCAAAAGTGGCTATTTACAGTCACTGATGATTTATATAGCAATA 539
 OY 1134 tataatccattatgcaagaactcaagaactccttgaaggagactctatagtatc 1193
 DB 540 TATTAATCTTATGTTAGCACTCAAGACATCCCTGCAAGGAGACTCTTATATGCTAT 599
 OY 1194 tgaagaagcttggaagcttcaatcattatgctgagctcgaatcctcactg 1253
 DB 600 TGAAGAAGCTGTGAAGCTTTCAGTTCCAATCTATATGAGTGGCTGCTGATTTCTACTG 659
 OY 1254 ccttcacaccttggttgaacatctgcaagaagctcctgctcggagagctggaatt 1313
 DB 660 CTCTTCACCTTTGGTTAAACATATTTGGCAGAGCTCCCTGCTCTGCGGAGCCGTGAAT 719
 OY 1314 ctcaagaagcttggaatgcaaaagctggaagattacaggagaaatggaatggaattgc 1373
 DB 720 CTCAAAAGATTTGGTGAAGTCAAAAGCGTTGGACATTTATTTGAGAAATGTCGAATATGCG 779
 OY 1374 tgcataaagtatgagcttcgacatatactcctcgtgcttgcgacagaagaatac 1433
 DB 780 TGTTCACAAATGATGATGTTCCGACATGATATCTTCCTGCTGCGCATCAAGATACCAA 839
 OY 1434 gaaacacgcatatcattgcttccatagctcgcagcttccatgagatagatgcagc 1493
 DB 840 AGTACCCCGCATTTATCATATGCTTTCTTACTGCTGAGCTTTTCATGATGATGATGATG 899
 OY 1494 agtctctgctcgtcctcctcaagctatgagcttcttggagatagatgctcagctc 1553
 DB 900 AGTCTCTGCTGCTGCTCTTCATATGCTGCTTCATGAGGAAATTTATGTTCACTGCTCTT 959
 OY 1554 gctctcacaacatcctcaagaagaagcttgcctcaagcggagaggaatgactc 1613
 DB 960 GGTCTTATTCACAACTTTTACAAAGAGGTTGGCTTCATGATGATGATGATGATGATG 1019
 OY 1614 ctggttcatcttctgcaatcttcggaacacgagatgctgctccttcttcttcttcc 1673
 DB 1020 TGGTTCAGCTCTTTCATATTTTCGACAAACCGATGTTGGGCTCTTATTTACATGAC 1079
 OY 1674 gatgaacgcaaaagagatgctatgctga 1703
 DB 1080 GATGAACCGCAAGATTCATCTCTGCAAA 1109

RESULT 10
 AFI29003
 LOCUS
 DEFINITION Nicotiana tabacum diacylglycerol acyl-CoA acyltransferase (DAGAT)
 ACCESSION AFI29003
 VERSION AFI29003
 KEYWORDS
 SOURCE common tobacco
 ORGANISM Nicotiana tabacum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 1 (bases 1 to 2099)
 Bouvier-Nave, P., Benveniste, P., Oelkers, P., Sturley, S.L. and
 Schaller, H.
 Expression in yeast and tobacco of plant cDNAs encoding acyl
 CoA:diacylglycerol acyltransferase
 Eur. J. Biochem. 267 (1), 85-96 (2000)
 2 (bases 1 to 2099)
 Benveniste, P.
 Direct Submission
 Submitted (17-FEB-1999) Plant Molecular Biology Institute,
 C.N.R.S., 28 rue Goethe, Strasbourg 67083, France
 FEATURES
 source
 /organism="Nicotiana tabacum"

OY 376 gataataacggtggtgcatataacgagtgatggaagagcgcggaagaagaaga 435
 DB 469 GAGGACGACACGAAATGATATGTTGGAAATGTTGAGGAAAGCAAAATCAACGAAACAA 528
 OY 436 aacgcacatgctagcttgaatcgcagctggttcagcgcacacgagagcgagagag 495
 DB 529 ACGACGCGCTTAAATTTGCTTACAGGGGCTGCGACACAGCTACCGCGGATCAAGAG 588
 OY 496 agtccactgagctcgagcagcaatcctcaacagagcagcagctatcaccctcgt 555
 DB 589 AGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 648
 OY 556 gtatgcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 615
 DB 649 GTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 708
 OY 616 tggatcagaagaagatctcgttgaatcagaatcagctgagagattggcgcttctc 675
 DB 709 CTTTATATAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 768
 OY 676 atgtgtgtatatacccttctgacttcttcttcttcttcttcttcttcttcttctt 735
 DB 769 ATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 828
 OY 736 ctgaagaatcatalatagaacactgtgtatccttcttcttcttcttcttcttcttctt 795
 DB 829 CAGCAGAGGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 888
 OY 796 gagcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 855
 DB 889 GCAATTTTGTATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 948
 OY 856 acttgatgctccttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 915
 DB 949 ATATTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1008
 OY 916 tatgaataagaatccctagcgaatgcaagctgataag-----caactcgaagtc 966
 DB 1009 CATGATATGACACAGCTGCAAGAGCTTACGACAGAGATGAAACTTCAAGATGGGATTT 1068
 OY 967 tcttactagcttgaagagcttgatgcatatctatgctgctccacatgcttat 1026

BASE COUNT 543 a 405 c 485 g 666 t
 ORIGIN

Query Match 36.68; Score 697; DB 8; Length 2099;
 Best Local Similarity 71.55; Pred. No. 3.4e-168;
 Matches 948; Conservative 0; Mismatches 365; Indels 12; Gaps 2;

/cultivar="xanthi sh6"
 /db_xref="taxon:4097"
 /tissue_type="calli derived from leaf protoplasts"
 1..2099
 /gene="DAGAT"
 208..1806
 /function="catalyzes the acylation of diacylglycerol by
 oleoyl-CoA to give triacylglycerol"
 /note="similar to acyl-CoA cholesterol acyltransferase"
 /product="diacylglycerol acyl-CoA acyltransferase"
 /protein_id="AAP19345.1"
 /db_xref="GI:6625653"
 /translation="MVMPLPESVEMTTTTTSGIENLNSDLNHSVRRRSGNGFEAA
 SAINSDAMSRDRDNYGSGAGLETENERSKSYGESSDVRREDDRNRYNANGESK
 ENIMKGLIRAGFWSSKLRDPLMLCISLIDILFLAFLVEXKLAQORHLTERAV
 TLHTTTTALILYPLVILIGDSEAFLEGVILMLVACTVMKILVSTAHNDRRLAKS
 TKDETSQDPSYDSEKSLAYFMVAPLTCYOLSYPTPCIRKGVAKOPIFLVETG
 LAGFLIEQYINPIYQNSQHPKGLVLAIRVYLKSLVPIYVLMWCFEFLMLNL
 AELLCEGDFREFYKDMNNAKTIDEVYMRPVRKMMVRIYEPCLRNCPKQVALILA
 ELVSAVFHELCLIAVPCRLPKMMAFMGIMFOVPLVILNPLONKFOSSMVGNMFWCFP
 CILGPMCVILYLYHVDVNRKRSAR"

Db 1069 TCTTATGATGTTAGCTTCAAGATTGGCTTACTCATGTTGGCCCAACATTTATGTTAT 1128
 QY 1027 CAGCAAGATGATCCAGTCTGCTGATGATACGAGAGGTTGGTGGTCTGCAATTCGCA 1086
 Db 1129 CAGCTTATGATCCCACTGCTGATGATGATGATGATGATGATGATGATGATGATG 1188
 QY 1087 AACTGTCATGATCCAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 1146
 Db 1189 AAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1248
 QY 1147 GTCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1206
 Db 1249 GTCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1308
 QY 1207 AAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1266
 Db 1309 AAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1368
 QY 1267 TGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1326
 Db 1369 TGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1428
 QY 1327 TGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1386
 Db 1429 TGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1488
 QY 1387 TGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1446
 Db 1489 TGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1548
 QY 1447 TGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1506
 Db 1549 TGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1608
 QY 1507 TGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1566
 Db 1609 TGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1668
 QY 1567 TGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1626
 Db 1669 TGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1728
 QY 1624 TGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1686
 Db 1729 TGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1788
 QY 1684 TGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1746
 Db 1789 TGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1808

FEATURES
 source
 gene
 CDS
 Seedun-dong, Gwonsun-gu, Suwon 441-707, Korea
 Location/Qualifiers
 1..1964
 /organism="Perilla frutescens"
 /cultivar="Okdong"
 /db_xref="taxon:48386"
 1..1964
 /gene="Dgat1"
 69..1673
 /gene="Dgat1"
 /note="Dgat1"
 /codon_start=1
 /product="diacylglycerol acyltransferase"
 /protein_id="AAC33696.1"
 /db_xref="GI:10803053"
 /translation="MAILDSPEILIDTSSADNGAAHHTTLRRROSASVPLINDSDS
 /transl_except="NSLEASINSENVANDANTLEIRGCAVESSENEQSYKGEKAEKENGSTNEN
 NSLEASINSENVANDANTLEIRGCAVESSENEQSYKGEKAEKENGSTNEN
 GTDMAVAFTEHPAPRPHARKRESPLSDAIFQSHAGLFNLCIVLVAVNSRLIEN
 LMKYGLIKSGFWSSISLRDWPILMCCISLPVALASFLVEKLVNLIPWAVFL
 HTYITVELLEPVPVILKDSAVLSGVTLCYOPSPYPRACIRKGMVROLIKVIF
 KPMASRWNLIDYAYDVSFKSLATFVMAPTLCYOPSPYPRACIRKGMVROLIKVIF
 TGLMGFIROYINPIVONSQHPKGNLIVATERYKLSYNNLVYVIMCFEPEHMLN
 ILAELIFEDREYEDKMMNARTVEYEMMMNPMVHKMVRHIIYCPCLONGIPKIVLV
 IATLVSAIRFHEILCAVPCOIFKFMFSGIMLVPLVIVTNILOEKFKNSVGMFMFC
 FCFIEQPMCVILYVHDLNKRKASAR"
 BASE COUNT 522 a 395 c 461 g 586 t
 ORIGIN
 Query Match 36.3%; Score 692; DB 8; Length 1964;
 Best Local Similarity 66.1%; Pred. No. 6; 5e-167;
 Matches 1067; Conservative 0; Mismatches 505; Indels 42; Gaps 3;
 QY 116 TGTTCGCAAAATTTCCGTGCGATGCGGATCTGGATCGCGGAGATCTTGACAGCA 175
 Db 46 TGTTCGCAAAATTTCCGTGCGATGCGGATCTGGATCGCGGAGATCTTGACAGCA 105
 QY 176 CGGAGAGCGTGCAG 235
 Db 106 CGTTCGCAAAATTTCCGTGCGATGCGGATCTGGATCGCGGAGATCTTGACAGCA 165
 QY 236 CGGAGAGCGTGCAG 277
 Db 166 CGGAGAGCGTGCAG 225
 QY 278 ATTCCTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 337
 Db 226 TCAATGATTCGCAAAATTTCCGTGCGATGCGGATCTGGATCGCGGAGATCTTGACAGCA 285
 QY 338 TTAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 397
 Db 286 CCGTTCGCAAAATTTCCGTGCGATGCGGATCTGGATCGCGGAGATCTTGACAGCA 345
 QY 398 ATAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 457
 Db 346 AGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 405
 QY 458 ATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 517
 Db 406 TCAAGCGCGGCG 465
 QY 518 TCTTCAAG 577
 Db 466 TCTTCAAG 525
 QY 578 ACAGTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 637
 Db 526 ATACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 585
 QY 638 GTTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 697
 Db 586 GTTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 645

1

```

/organism="Arabidopsis thaliana"
/cultivar="Columbia"

```


misc_feature

<1..1671
/note="Sequence from clone F27F23"
complement(join(<1311..1511,1608..2357,2445..2873,
3102..3203,3293..3567,3957..4070,4242..4380,4652..4824,
5200..>5422))
/gene="At2g19410"
.55422)
complement(<1311..
/gene="At2g19410"
/note="F27F23.21; contains a protein kinase domain profile
(PDOC00100)"
complement((join(1311..1511,1608..2357,2445..2873,
3102..3203,3293..3567,3957..4070,4242..4380,4652..4824,
5200..5422))
/gene="At2g19410"
/codon_start=1
/product="putative protein kinase"
/protein_id="AADI0166.2"
/db_xref="GI:6598498"

CDS

/translation="MYVMLTQEMSGGCGKPAEKGOLFVAVAVKLGIDGKLGAGSRRR
VRVAADILAKRAKFVMYHIVIPITSISPIINLIIMFTMMWVTAGDRLPYAEVEEV
IEMVYRVKKKEIEIVPRPLMKCKSTRXYRSRRGTGCVPLRLKYAPETCEIA
IVCDRIITYSMDLIRNEECTSPHAATADHLROMASFFHLSPTLDIPROSTEA
GTRSAASENLTREFALSLTCNKRPETPOSSKASSATTPEIFRRRGSDILPOLNSDEK
TCTPQSASVENTVSSEHRSDSRPETSRSKAYIEEEVERKNELOSIFYFKQACE
ELFSFONORIAEVNLRTYLEKKKVIDDLCTDRHYKRITTELYVTMEGPSPEKVC
AREPCORIAEVNLRTYLEKKKVIDDLCTDRHYKRITTELYVTMEGPSPEKVC
GGCYGVCCSIDSTPAAVKVVRLDTPEPKKOEFLEKVEVLSQLRHPHYVLIIIGCPNG
CLVITYENGLSILEYIFHKRKPLPLPFIFRVIIEVACGIAFLHSKPREPIYHRDK
PINYLNNRYTSKIADVGTLAKTMDVADPNWTYRNLSVLAGTTHYIDPEHRTGTRP
KSDIAFGIITLLDILTANPGCIYAVENAYKKTLEMJLMDKSSTDMPLAETHELAKI
GSCAFECFRDRPDLSKEYIIVLIKRVLTANSKYKKGSNLRAPASHFCPILEIMEE
PVY"

misc_feature

1672..>92721
/note="Sequence from clone F3p11"
join(<811..8129,8632..8742,9989..>10293)
/gene="At2g19420"
<811..>10293
/gene="At2g19420"
/note="F3p11.2; predicted by gensecan"
join(811..8129,8632..8742,9989..10293)
/gene="At2g19420"
/note="hypothetical protein"
/codon_start=1
/protein_id="AADI0141.1"
/db_xref="GI:4191772"
/translation="MHEKVPRKTYDSELNYGYLAIFYNGYGLDYRFYTKIENDICNSLV
WNTEMTRKSRRKRIIRINYRPRKAPKKNKPPTFEQLMDPEFSDDVLTLSGFDDG
GYIPLPKRSLINLPDSFDELQTMHVFEYLREVGKLMEG"

gene

8188..8269
/rpt_family="POLY_A"
/rpt_region
8824..8918
/rpt_family="(TA)"
join(<11627..11952,12040..12116,12198..12243,12371..12431,
12520..12573,12664..12744,12836..12875,13148..13248,
13421..13531,14061..>14150)
/gene="At2g19430"
/note="At2g19430"
/gene="F3p11.3; contains a Trp-Asp (WD-40) repeat
signature"
join(11627..11952,12040..12116,12198..12243,12371..12431,
12520..12573,12664..12744,12836..12875,13148..13248,
13421..13531,14061..14150)
/gene="At2g19430"
/codon_start=1
/product="putative WD-40 repeat protein"
/protein_id="AADI0142.1"
/db_xref="GI:4191773"

CDS

/translation="MGYDATNMNDEYRESILKERIETRYTFRTAMAPPARISNDPA
FYVASGQTIALPHSLINSOSAFSGSKODVWAVEPERVVARAHGPAYVDKPYGED
EDALLSLGCDGRRCRWKMKREFASDVSLHLKEMHLPLELILNPQHKGPMGALSMP

[illegible]

QY 241 tcttctaagcagctctctctcgtggtccgataataatcccttcgagatgattgta 300
 Db 22948 tcttctaagcagctctctcgtggtccgataataatcccttcgagatgattgta 300
 QY 301 gctccggcgcagcgttaaggatcagatgattcgttgaacagcagcagcagc 23007
 Db 23008 gctccggcgcagcgttaaggatcagatgattcgttgaacagcagcagcagc 23007
 QY 361 gccaatgtgctgagagataataacggtggtgagataataacggtggtgagata 23067
 Db 23068 gccaatgtgctgagagataataacggtggtgagataataacggtggtgagata 23067
 QY 421 ggaagaag 23127
 Db 23128 ggaagaag 23127
 QY 481 cggagggcggag 23187
 Db 23188 cggagggcggag 23187
 RESULT 13
 AF221132
 LOCUS 1497 bp mRNA linear INV 14-JUL-2000
 DEFINITION Caenorhabditis elegans diacylglycerol acyltransferase mRNA,
 complete cds.
 ACCESSION AF221132
 VERSION AF221132.1 GI:9049537
 KEYWORDS
 SOURCE
 ORGANISM
 Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 1497)
 Bouvier-Nave, P., Benveniste, P., Noirel, A. and Schaller, H.
 Expression in yeast of a cDNA from Caenorhabditis elegans encoding
 a diacylglycerol acyltransferase (DGAT)
 2. (bases 1 to 1497)
 Bouvier-Nave, P. and Benveniste, P.
 Direct Submission
 Submitted (05-JAN-2000) Plant Molecular Biology Institute, CNRS, 28
 rue Goethe, Strasbourg 67083, France
 Location/Qualifiers
 1. 1497
 /organism="Caenorhabditis elegans"
 /strain="CB1489 him-8(e1489)"
 /db_xref="dbEST:c47874"
 /db_xref="taxon:6239"
 /note="Original clone yk453a2 provided by Yui Kohara"
 1. 1497
 /EC-number="2.3.1.20"
 /function="terminal enzyme during triacylglycerol
 biosynthesis"
 /note="DGAT: functional diacylglycerol acyltransferase
 when expressed in yeast"
 /codon_start=1
 /product="diacylglycerol acyltransferase"
 /protein_id="AA082410.1"
 /db_xref="GI:9049538"
 /translation="MOMROQTGRRRRPPSTNSGSLSSRRSSPAONGNSRRSSSMR
 GPEKVVHTADSLFSTSSGWNFRGPFNLSILLVLSNKRVALENYIKYLITPIQ
 WISTVEHHYSIWPNLALILGNTIOISFSGEKLLEGGNGENFAVFTSLVTA
 HLIIPVVVTLTHKKNPLMPLSVVMGYVIEALKFTISGHVNWADARARKITELKOV
 TDLAKTCDKPOFNDIKELSMHQAQVPAKLTLSIYFMAAPLICEPEPPLIR
 IRKHLIKRTVELFPLNLALAVQWVFTVNSMKPLISEMYSCLERILKIAIPN
 HLWLIFPTTFPSFLNLALVQWVFTVNSMKPLISEMYSCLERILKIAIPN
 IYSPMRNKSMSAFVEFVSFAFEHETLVSPDKIFPLMSYGMAGGQIPLSITDK
 VVGRGRTNIIIVSLIVGQPLAIVMGHDWYILNCSAVAVQNVQTVG"
 BASE COUNT 402 a 311 c 308 g 476 t
 ORIGIN

Query Match 8.1%; Score 154.2; DB 3; Length 1497;
 Best Local Similarity 56.7%; Pred. No. 1.1e-28;
 Matches 325; Conservative 0; Mismatches 243; Indels 5; Gaps 2;

QY 997 tattcaatgctgcctccacatgtgtatcaagcaagatccacgcttctcatgata 1056
 Db 772 taccatgctgcctccacatgtgtatcaagcaagatccacgcttctcatgata 1056
 QY 1057 cggagggcttggtggtcgcctcgaatttgcgaacatgctcgaatccacgcttctcatgata 831
 Db 832 cggagggcttggtggtcgcctcgaatttgcgaacatgctcgaatccacgcttctcatgata 831
 QY 1117 ttataatagaacaataataataacccatgtgtcgaagcaactcgaatccacgcttctcatgata 891
 Db 892 gcaactgtttcaacatggtggtgttcgacgttcgcaaatgataaaccctttaaagtaa 951
 QY 1173 aggcacatctatagctatgataagagagtgtaagcttccagcttccacatgata 1232
 Db 952 atggaatgcttgaatggtt-tggaacgactcttgaacatgcaatgcaatgcaatgcaat 1010
 QY 1233 gtagctcgaatgcttgaatgcttccacatgcttccacatgcttccacatgcttccacat 1292
 Db 1011 ctgactgttctggttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1070
 QY 1293 ctgactgttctggttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1352
 Db 1071 tgcatttgcgacatgcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1130
 QY 1353 ctgaggaatgcttgaatgcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1412
 Db 1131 ctggaatgcttgaatgcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1190
 QY 1413 ctgaggaatgcttgaatgcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1472
 Db 1191 gattgcttgaatgcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1250
 QY 1473 ctgaggaatgcttgaatgcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1532
 Db 1251 ctgaggaatgcttgaatgcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1310
 QY 1533 gattgcttgaatgcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1565
 Db 1311 aatgattggaacaaattcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1343

RESULT 14
 LOCUS CEH19N07 11966 bp DNA linear INV 11-DEC-2001
 DEFINITION Caenorhabditis elegans cosmid H19N07, complete sequence.
 ACCESSION Z92835
 VERSION Z92835.1 GI:2414214
 KEYWORDS HTG.
 SOURCE
 ORGANISM
 Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (sites)
 none.
 REFERENCE 1
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 99069613
 REMARK The C.elegans Sequencing Consortium.
 2 (bases 1 to 11966)
 AUTHOR DOBSON, R.
 TITLE
 JOURNAL
 Submitted (09-MAR-1997) Nematode Sequencing Project, Sanger
 Institute, Hinxton, Cambridge CB10 1SA, England and Department of
 Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
 jes@sanger.ac.uk or rwenematode.wustl.edu
 On Sep 20, 1997 this sequence version replaced gi:1877112.
 Coding sequences below are predicted from computer analysis, using
 predictions from GeneFINDER (P. Green, U. Washington), and other

COMMENT

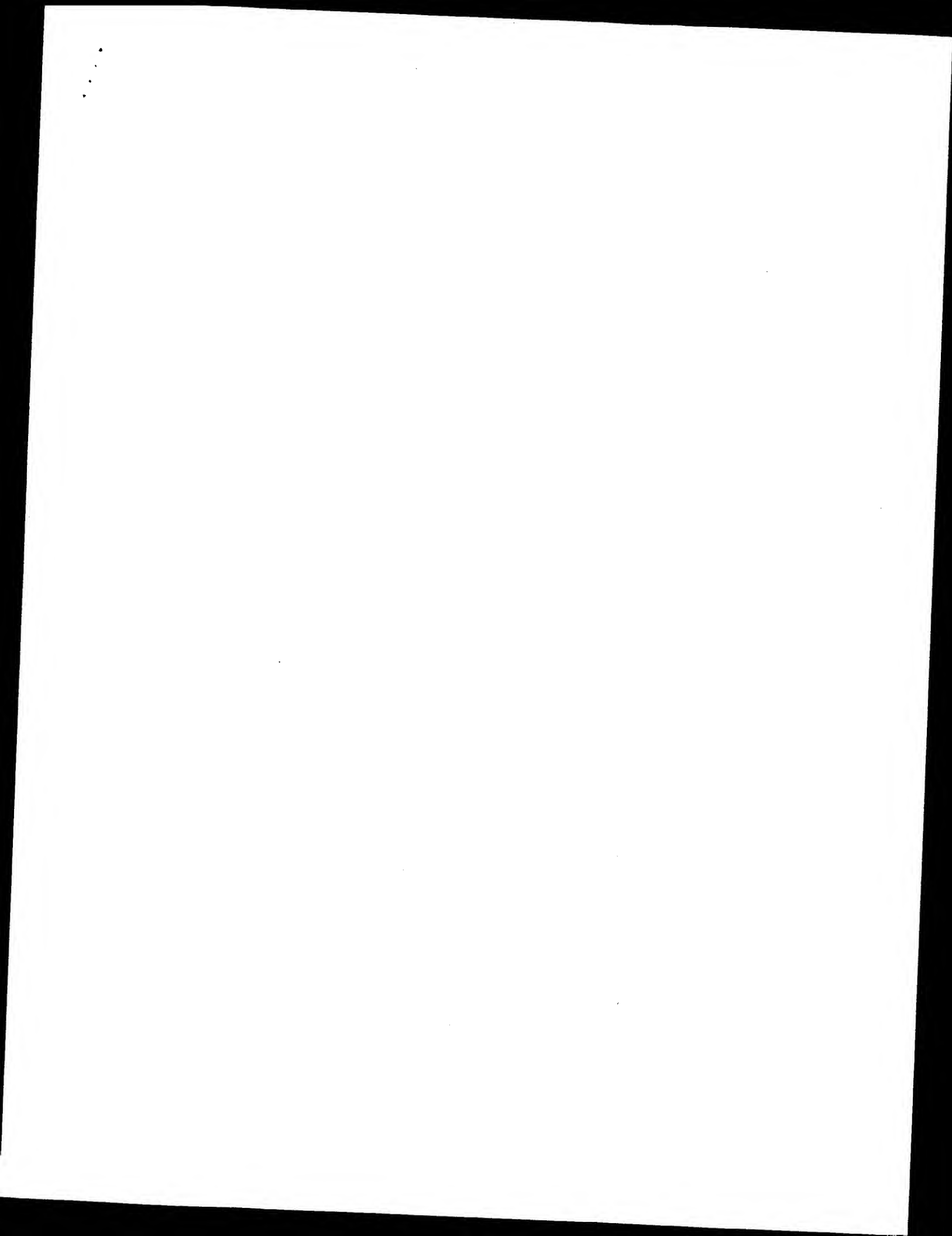
Mon May 13 10:17:43 2002

us-09-623-514a-1.rge

Page 19

QY 1234 tggctctgcatggttctactgctcttccaccttggthaaacatatggcagaagcttctc 1293
1128 TGGCTTATCTTCTTCTATGTGGTTTTCACCTCTCAATGCTGTGGCAGAGCTTCTG 1187
QY 1294 tgcctcgggagatcgtgaattctacaagaattggtggaatgcanaaagtglyggaaltac 1353
1188 CAGTTTGGAGACCGCGAGTTCACAGAGATGGTGAAATGCTGAGTCTGTCACTTACTTT 1247
QY 1354 tggagaatggtggaatatgctctgttcaaaatggaatggtggaatataacttccgltgc 1413
1248 TGGCAGAACTGGAAATATCCCGTCACAAAGTGTGCATCAGACACTTACAGGCTATG 1307
QY 1414 tgcgcagcagaagataccaagaacactcggcattatcatctgcttccctagctcgtcagtc 1473
1308 CTCAGACATGGCAGCAGCAAAATGGGTGGCCAGGACAGAGATATTTTTCACCTCAGGCTTC 1367
QY 1474 ttccatggcctatgcatcgcagtccttgcgtctcttcaagctatgggcttcttctg 1533
1368 TTCATGAGTACCTATGAGGCTTCCCTGGGATGTCGCTCTGGGCAATTCACAGGC 1427
QY 1534 ataatgttcaaggcttggcttgcatt 1562
1428 ATGATGCTCAGGTCCTCAGCTGGCTGGAT 1456
Db

Search completed: May 13, 2002, 07:18:25
Job time: 17074 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 03:31:56 ; Search time 721.3 Seconds
(without alignments)
4532.102 Million cell updates/sec

Title: US-09-623-514A-1

Perfect score: 1904

Sequence: 1 attcttagctctctccttc.....ttatgtgtgaagcgcgcg 1904

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1904	100.0	1904	21	AAA51482	A. thaliana diacyl
2	1883.2	98.9	1888	21	AAA48932	Arabidopsis diacyl
3	1839.4	96.6	1942	21	AAA88835	Arabidopsis acyl C
4	1839.4	96.6	1942	21	AAZ45371	Acyl-CoA:cholester
5	1839.4	96.6	1942	22	AA501106	Arabidopsis thalia
6	1813	95.2	1965	21	AAA51484	A. thaliana AS11 d
7	719.8	37.8	1942	21	AAA48839	Soybean diacylglyc
8	575.2	30.2	1975	21	AAA48942	wheat diacylglycer
9	542.4	28.5	1587	21	AAA48938	Rice diacylglycer

10	528	27.7	5193	21	AAA51483	A. thaliana diacyl
11	528	27.7	5339	21	AAA51485	A. thaliana AS11 d
12	424	22.3	1281	21	AAA48933	Corn diacylglycer
13	379.4	19.9	1559	21	AAA48935	Corn diacylglycer
14	355.2	18.7	901	21	AAA48936	Corn diacylglycer
15	336.4	17.7	629	21	AAA51486	EST with homology
16	336.4	17.7	629	21	AAZ49453	A. thaliana diacyl
17	239.8	12.6	978	21	AAA48934	Corn diacylglycer
18	141.8	7.4	1650	21	AAZ49452	Mouse diacylglycer
19	141.2	7.4	1521	19	AAV01533	Human acylCoenzyme
20	141.2	7.4	1976	21	AAZ76169	Human ACAT Related
21	139.6	7.3	993	22	ABA06413	Human acyl CoA:cho
22	139.6	7.3	1895	21	AAZ45383	DNA encoding a pro
23	139.6	7.2	1766	21	AAZ45384	Rat acyl CoA:chole
24	137.6	7.2	1766	21	AAZ45385	Rat acyl CoA:chole
25	137.6	7.2	1766	22	AA501105	Rat sterol acyltra
26	137.6	7.2	275	21	AAA88840	Soybean acyl CoA:c
27	125.8	6.6	275	21	AAZ45374	Acyl-CoA:cholester
28	125.8	6.6	275	21	AAZ45374	Soybean sterol acy
29	125.8	6.6	275	22	AA501313	Human acylCoenzyme
30	124.6	6.5	983	19	AAV01539	Soybean acyl CoA:c
31	124.6	6.4	234	21	AAA88839	Human acyl CoA:cho
32	122.6	6.4	234	21	AAZ45372	Acyl-CoA:cholester
33	122.6	6.4	234	22	AA501311	Soybean sterol acy
34	122.6	6.4	254	21	AAA48941	Rice diacylglycer
35	121.6	6.0	466	21	AAA48941	Wheat diacylglycer
36	115	6.0	366	22	AAZ45377	Rice diacylglycer
37	109.4	5.7	254	21	AAZ45377	Maize acyl CoA:cho
38	109.4	5.7	254	21	AAZ45377	Acyl-CoA:cholester
39	109.4	5.7	254	22	AA501316	Maize sterol acylt
40	108.2	5.7	267	21	AAA88841	Soybean acyl CoA:c
41	108.2	5.7	267	21	AAZ45373	Acyl-CoA:cholester
42	108.2	5.7	267	22	AA501312	Soybean sterol acy
43	105	5.5	1132	20	AAZ30335	DNA encoding a hum
44	103.6	5.4	262	21	AAA88848	Maize acyl CoA:cho
45	103.6	5.4	262	21	AAZ45378	Acyl-CoA:cholester

ALIGNMENTS

RESULT 1	AAA51482	standard: cDNA: 1904 BP.
ID	AAA51482	
XX	AAA51482:	
AC	09-OCT-2000 (first entry)	
XX		
XX	A. thaliana diacylglycerol acyltransferase cDNA.	
DE		
XX	DEAT: diacylglycerol acyltransferase; seed oil; fatty acid synthesis;	
KW	size: weight; carbon flux: ss.	
KW		
XX	Arabidopsis thaliana.	
OS		
XX		
XX	Key	Location/Qualifiers
XX	CDS	139..1701
XX		/*tag= a
XX		/product= diacylglycerol_acyltransferase
XX		
XX	WO200036114-A1.	
XX	22-JUN-2000.	
XX		
XX	16-DEC-1999;	99WO-CAN01202.
XX		
XX	17-DEC-1998;	98US-0112812.
XX		
XX	(CANVA) NAT RES COUNCIL CANADA.	
XX		
XX	Zou J, Taylor DC, Wei Y, Juko CC;	
XX		

QY 1149 caggaaactcaaaagcatctcttgaaagcgatctctataatgtaataagagttgaa 1208
 Db 1141 caggaaactcaaaagcatctcttgaaagcgatctctataatgtaataagagttgaa 1200
 QY 1209 gcttcaatccaataatataatgtaataagagttgaa 1208
 Db 1201 gcttcaatccaataatataatgtaataagagttgaa 1200
 QY 1269 gtttaacatattgcaagagctctctgctgaggaatcgttaataagagttgaa 1260
 Db 1261 gtttaacatattgcaagagctctctgctgaggaatcgttaataagagttgaa 1260
 QY 1329 gaatgcaaaaagtgaggaatcgttaataagagttgaa 1320
 Db 1321 gaatgcaaaaagtgaggaatcgttaataagagttgaa 1320
 QY 1389 gtttgcacataatactctccgctgctgaggaatcgttaataagagttgaa 1380
 Db 1381 gtttgcacataatactctccgctgctgaggaatcgttaataagagttgaa 1380
 QY 1449 cattgcttccagctctctgaggaatcgttaataagagttgaa 1440
 Db 1441 cattgcttccagctctctgaggaatcgttaataagagttgaa 1440
 QY 1509 cttcaagctatgagctctctgaggaatcgttaataagagttgaa 1500
 Db 1501 cttcaagctatgagctctctgaggaatcgttaataagagttgaa 1500
 QY 1569 ctatcacaaggaaggttgaggaatcgttaataagagttgaa 1560
 Db 1561 ctatcacaaggaaggttgaggaatcgttaataagagttgaa 1560
 QY 1629 catttgcgaacacaggttgaggaatcgttaataagagttgaa 1620
 Db 1621 catttgcgaacacaggttgaggaatcgttaataagagttgaa 1620
 QY 1689 atcgagtgatgaaacacgttgaggaatcgttaataagagttgaa 1680
 Db 1681 atcgagtgatgaaacacgttgaggaatcgttaataagagttgaa 1680
 QY 1749 ggaatcgcgttgaggttgaggaatcgttaataagagttgaa 1740
 Db 1741 ggaatcgcgttgaggttgaggaatcgttaataagagttgaa 1740
 QY 1809 tgaagaagaagaagaatgaggttgaggaatcgttaataagagttgaa 1800
 Db 1801 tgaagaagaagaagaatgaggttgaggaatcgttaataagagttgaa 1800
 QY 1869 acccggttgaggttgaggaatcgttaataagagttgaa 1860
 Db 1861 acccggttgaggttgaggaatcgttaataagagttgaa 1860

RESULT 3
 ID AAA88835 standard; DNA; 1942 BP.
 AC AAA88835;
 DT 19-FEB-2001 (first entry)
 XX Arabidopsis acyl CoA:cholesterol acyltransferase DNA.
 KW Acyl CoA:cholesterol acyltransferase; ACAT; transgenic plant;
 KW sterol; tocopherol; phytosterol; phytostanol; anticholesterolemic;
 XX hypolipemic; ss.
 OS Arabidopsis thaliana.
 PH Key Location/Qualifiers
 FT CDS 237..1799
 FT /*tag= a

FT /EC_number= "2.3.1.26"
 XX MO200061771-A2.
 PN 19-OCT-2000.
 XX 12-APR-2000; 2000WO-US09696.
 XX 12-APR-1999; 99US-0128995.
 XX (MONS) MONSANTO CO.
 PA Venkatramesh M, Corbin DR, Bhat GB, Boddupalli SS, Grebenok RJ;
 PI Kishore GM, Lardizabal KD, Lassner MW, Rangwala SH, Karunanandaa B;
 DR WPI: 2000-66516/64.
 DR P-PSDB: AAB19740.
 PS Genetically engineering the biosynthetic pathways in plants involved in
 the accumulation of sterol compounds and tocopherol to produce
 compounds for lowering the level of low density lipoprotein cholesterol
 in blood serum.
 PS Disclosure: Page 57-58; 16pp; English.
 CC The present sequence is that of Arabidopsis thaliana DNA encoding
 CC acyl CoA:cholesterol acyltransferase (ACAT, see AAB19740). Sterol
 CC O-acyltransferases such as ACAT catalyse the formation of
 CC cholesterol esters from cholesterol and long chain fatty acids.
 CC Recombinant constructs of the invention are used to alter the
 CC biosynthesis and accumulation of sterols and tocopherols in
 CC transgenic plants. Seeds of such plants may contain elevated
 CC levels of sitosterol and/or its esters, and alpha-tocopherol, and
 CC reduced levels of campesterol and campestanol and their esters.
 CC The seeds may also contain the novel sterol brassicatanol. Oil
 CC compositions from the seeds can be used in food and pharmaceutical
 CC compositions to lower levels of low density lipoprotein cholesterol
 CC in blood serum. ACAT enzymes can be used in the present invention
 CC to produce elevated levels of phytosterol and/or phytostanol esters.
 CC XX
 SQ Sequence 1942 BP; 461 A; 421 C; 425 G; 635 T; 0 other;

Query Match 96.6%; Score 1839.4; DB 21; Length 1942;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1840; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atttctagctctctctcaatccgctctctccctcattagattctgttccctt 60
 Db 99 atttctagctctctctcaatccgctctctccctcattagattctgttccctt 60
 QY 61 tcaattctctgcatgctctctctcgaagcctcttccctccatagattctgttccctt 158
 Db 159 tcaattctctgcatgctctctctcgaagcctcttccctccatagattctgttccctt 120
 QY 121 cgtcaaacgcttctcgaatgagcatttgagatctcgtcgtgcttaacaggtgagag 180
 Db 219 cgtcaaacgcttctcgaatgagcatttgagatctcgtcgtgcttaacaggtgagag 180
 QY 181 aacggtgaggaagctcgtcgaatcttgatagcttgcgaaggaatcgaatcgaatcgaat 240
 Db 279 aacggtgaggaagctcgtcgaatcttgatagcttgcgaaggaatcgaatcgaatcgaat 240
 QY 241 tcttcaacgagctctctctctcgtcgtcgaatataatctcctcgaatgatttga 300
 Db 339 tcttcaacgagctctctctctcgtcgtcgaatataatctcctcgaatgatttga 300
 QY 301 gctcccgccgaagcttgaagatgagatcgttcttcaacaggaatcgaatcgaatcgaat 398
 Db 399 gctcccgccgaagcttgaagatgagatcgttcttcaacaggaatcgaatcgaatcgaat 360
 QY 361 gccaatctgcccgaaggaataacaggtggtgcgaataacaggtggtgagagggcgc 420
 Db 458 gccaatctgcccgaaggaataacaggtggtgcgaataacaggtggtgagagggcgc 420

XX The present sequence encodes an acyl-CoA:cholesterol acyltransferase (ACAT) related protein. The ACAT-like protein is active in the formation of a sterol, ester and/or triacylglycerol from a fatty acyl-CoA and sterol and/or diacylglycerol substrate. The DNA can be used for modifying the lipid composition of plant cells. The ACAT-like protein has diacylglycerol acyltransferase (DGAT) activity, and so the synthesis of triacylglycerides can be suppressed or increased using the DNA. The protein can be used to produce plant oils with a modified triacylglyceride content. The products can also be used to identify antagonists and agonists of DGAT activity. Such agonists and antagonists are particularly useful in treating or ameliorating diseases associated with DGAT activity, including or associated with altered cellular diacylglycerol concentration or PKC activity, including cancer, diabetes, cardiopulmonary diseases e.g. heart failure, atherosclerosis, adipocytosis, leukaemia, skin carcinoma, fibroadenoma, metabolic disorders, obesity, diseases associated with abnormal lipid metabolism, and diseases associated with abnormal fat absorption, lipoprotein secretion and adipogenesis.

Sequence 1942 BP; 461 A; 421 C; 425 G; 635 T; 0 other:

Query Match 96.6%; Score 1839.4; DB 21; Length 1942;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1840; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 attcttagcttcttcttccatccgcttcttcccttccatctagatctgttctctt 60
DB attcttagcttcttcttccatccgcttcttcccttccatctagatctgttctctt 158
QY 61 tcaattctcttgcatgtcttctcgtatctctctgacgctcttcttccgacgtctt 120
DB 159 tcaattctcttgcatgtcttctcgtatctctctgacgctcttcttccgacgtctt 218
QY 121 cgtcaaacgcttcttgcgaatttgatctctgctgacgtcttcttccgacgtctt 180
DB 219 cgtcaaacgcttcttgcgaatttgatctctgctgacgtcttcttccgacgtctt 278
QY 181 aacgctgacgagagctctgcatcttcatagctctgacgaggaattgagatcgat 240
DB 279 aacgctgacgagagctctgcatcttcatagctctgacgaggaattgagatcgat 338
QY 241 tcttcaacgacactctctctctgtgttccgataaattcttcccttgcgaatgttga 300
DB 339 tcttcaacgacactctctctctgtgttccgataaattcttcccttgcgaatgttga 398
QY 301 gctcccgacgactttagagatcgatgattcgtgttaacgataagctcaaggaa 360
DB 399 gctcccgacgactttagagatcgatgattcgtgttaacgataagctcaaggaa 458
QY 361 gccaatctgacgagataataacggtgtgacgataaagctggtgaagagcgagc 420
DB 459 gccaatctgacgagataataacggtgtgacgataaagctggtgaagagcgagc 518
QY 421 gggagagagagagaaacgacgtgtaacgtatcgatcgacgctggttccagat 480
DB 519 gggagagagagagaaacgacgtgtaacgtatcgatcgacgctggttccagat 578
QY 481 cggagagcagagagagatcacttagctcgcgacgaatttccaaacagagcagc 540
DB 579 cggagagcagagagagatcacttagctcgcgacgaatttccaaacagagcagc 638
QY 541 ttattcaacctgtgttagttagtcttattgtgtaacagtagacatcatatgaat 600
DB 639 ttattcaacctgtgttagttagtcttattgtgtaacagtagacatcatatgaat 698
QY 601 cttatgaagatgattgtgtttagatcaagagatcttctgtttagtcaagatcgctga 660
DB 699 cttatgaagatgattgtgtttagatcaagagatcttctgtttagtcaagatcgctga 758
QY 661 gattgacgcttcatgtgttatacccttgcgatttcttgcgtcccttaag 720

DB 759 gattgacgcttcatgtgttatacccttgcgatttcttgcgtcccttaag 818
QY 721 gttgagaattgtgacttcagaataatatacagaacctgtgtcatcttcttcaat 780
DB 819 gttgagaattgtgacttcagaataatatacagaacctgtgtcatcttcttcaat 878
QY 781 attacacacatgacagaggtttgtatccagtttaacgacccaatgaggtatctgt 840
DB 879 attacacacatgacagaggtttgtatccagtttaacgacccaatgaggtatctgt 938
QY 841 ttattataggtgacattttagatgcttcttcttcttcttcttcttcttcttctt 900
DB 939 ttattataggtgacattttagatgcttcttcttcttcttcttcttcttcttctt 998
QY 901 tatgtcatatcagctatgacataaagatccctagacacagcagctgaagcgaatct 960
DB 999 tatgtcatatcagctatgacataaagatccctagacacagcagctgaagcgaatct 1058
QY 961 gaagctccactacgattagactgaaagcttggacatattcaatgtgctccacattg 1020
DB 1059 gaagctccactacgattagactgaaagcttggacatattcaatgtgctccacattg 1118
QY 1021 tgttatcagcgaagttatccacgttctgcatgtatcaggaaggtgtgtgtcgtcaa 1080
DB 1119 tgttatcagcgaagttatccacgttctgcatgtatcaggaaggtgtgtgtcgtcaa 1178
QY 1081 ttgtcaaacctgtgcatatccacgattcaggtggaattataaagaacataataat 1140
DB 1179 ttgtcaaacctgtgcatatccacgattcaggtggaattataaagaacataataat 1238
QY 1141 ccatgtgacgaacatccaaacatcccttgaagacgacttccatattgtattgaaga 1200
DB 1239 ccatgtgacgaacatccaaacatcccttgaagacgacttccatattgtattgaaga 1298
QY 1201 ggtgtgaagcttccagttccaaattatgtgtgtcgtcgtatgttctacgtcttc 1260
DB 1299 ggtgtgaagcttccagttccaaattatgtgtgtcgtcgtatgttctacgtcttc 1358
QY 1261 caaccttggtaaacaatttggacagacgtctcgtctcgtggaatcgatcttca 1320
DB 1359 caaccttggtaaacaatttggacagacgtctcgtctcgtggaatcgatcttca 1418
QY 1321 gattgtgtgaatgtgaaatgtgtgagatcttgaagagatgtgagatgctcgtat 1380
DB 1419 gattgtgtgaatgtgaaatgtgtgagatcttgaagagatgtgagatgctcgtat 1478
QY 1381 aatgtgtgtgtgacatataatcttccgtgtgtgtgacgagataccaaagacact 1440
DB 1479 aatgtgtgtgtgacatataatcttccgtgtgtgtgacgagataccaaagacact 1538
QY 1441 gccattatcattgttcttctagctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1500
DB 1539 gccattatcattgttcttctagctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1598
QY 1501 tgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1560
DB 1599 tgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1658
QY 1561 atcaacaactatcaagaaagcttgcgcaacaggtgtggaacacagatccttgcgtc 1620
DB 1659 atcaacaactatcaagaaagcttgcgcaacaggtgtggaacacagatccttgcgtc 1718
QY 1621 atcttctgacatttggacacacagatgtgtgtcgtcgtcgtcgtcgtcgtcgtcgt 1680
DB 1719 atcttctgacatttggacacacagatgtgtgtcgtcgtcgtcgtcgtcgtcgtcgt 1778
QY 1681 cgaagaagatcgaatgtcgaacaacagcttcaaaaataaacttcttcaacaatcatg 1740
DB 1779 cgaagaagatcgaatgtcgaacaacagcttcaaaaataaacttcttcaacaatcatg 1838
QY 1741 gctcgtgtgtgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1800
DB 1839 gctcgtgtgtgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1898

Db 301 gctccgcgcagcttagagatcgatgtgattccgtgttaacagatgacgctcaggaaca 360
 Oy 361 gccaatcttgccgagagataaatacgtgtgtgagataaaccgtgtgagagagcgc 420
 Db 361 gccaatcttgccgagagataaatacgtgtgtgagataaaccgtgtgagagagcgc 420
 Oy 421 ggaagaaggaaggaaggaagcgcgagatgacgtgttaacgatalcagacgttcacatcat 480
 Db 421 ggaagaaggaaggaaggaagcgcgagatgacgtgttaacgatalcagacgttcacatcat 480
 Oy 481 cggagagcgagagagagatgacgtgttcgagacgacatcttcaaac----- 526
 Db 481 cggagagcgagagagagatgacgtgttcgagacgacatcttcaaacgagcgcgcga 540
 Oy 527 ----- 526
 Db 541 ttattcaacctgtgtgtagttagttcttaattgtglaaacagtagacatcaatcagaat 600
 Oy 527 -----agagcagatgcccagattatcaacctgtgtgtagttagttcttaattgtglaaac 579
 Db 601 cttaagagagcagatgcccagattatcaacctgtgtgtagttagttcttaattgtglaaac 660
 Oy 580 agtagagatcaatcagaatcttaagatgagtggtgtgagacgagatcttcg 639
 Db 661 agtagagatcaatcagaatcttaagatgagtggtgtgagacgagatcttcg 720
 Oy 640 tttagtcaagatcgctgagagatgagcgcgtttcagatggtgtgatactcccttcgac 699
 Db 721 tttagtcaagatcgctgagagatgagcgcgtttcagatggtgtgatactcccttcgac 780
 Oy 700 ttcccttggtgctgtttagcgtgtgagaaattgtagtctcagaataacatccagaacct 759
 Db 781 ttcccttggtgctgtttagcgtgtgagaaattgtagtctcagaataacatccagaacct 840
 Oy 760 gttgcatccttctcaatattatatacagatgacagagtttctgatalccagtttagc 819
 Db 841 gttgcatccttctcaatattatatacagatgacagagtttctgatalccagtttagc 900
 Oy 820 acccgaaggtgtgattctgctttttagcagtgatcagttgtagtctcagacat 879
 Db 901 acccgaaggtgtgattctgctttttagcagtgatcagttgtagtctcagacat 960
 Oy 880 gttggtgagaaattgtgttcttaagatgacatgacatgacatccagacat 939
 Db 961 gttggtgagaaattgtgttcttaagatgacatgacatgacatccagacat 1020
 Oy 940 gtagcgtgataagcgaatccctgaagttctcactagcttagctgagagagcttgacat 999
 Db 1021 gtagcgtgataagcgaatccctgaagttctcactagcttagctgagagagcttgacat 1080
 Oy 1000 ttcaatgctgctcccaatctgtgtatcagcgaagttatccagcttctgatalcagc 1059
 Db 1081 ttcaatgctgctcccaatctgtgtatcagcgaagttatccagcttctgatalcagc 1140
 Oy 1060 aagaggtgtggtgctgctgaatctgcaaaactgtgatalcagcgaagttatgagatt 1119
 Db 1141 aagaggtgtggtgctgctgaatctgcaaaactgtgatalcagcgaagttatgagatt 1200
 Oy 1120 ataataagacaataataataatctatgtcagaaactcaaacgacatcccttgaagagcat 1179
 Db 1201 ataataagacaataataataatctatgtcagaaactcaaacgacatcccttgaagagcat 1260
 Oy 1180 ctctcataatgataatgaagagtgtagagcttcaagcttcaaacatattatgtgtgctc 1239
 Db 1261 ctctcataatgataatgaagagtgtagagcttcaagcttcaaacatattatgtgtgctc 1320
 Oy 1240 tgcagtgctcactgcttctccacgttgcgttgaacatattatgtgtgctc 1299
 Db 1321 tgcagtgctcactgcttctccacgttgcgttgaacatattatgtgtgctc 1380
 Oy 1300 gggagatcgtaattctacaaagatgtgtgagacaaagctgtgtgagatctacgaga 1359
 Db 1381 gggagatcgtaattctacaaagatgtgtgagacaaagctgtgtgagatctacgaga 1440

Oy 1360 atgtgaaatagctgtgttcaataatgagatgtgttcagacatatacttccggtctgcgc 1419
 Db 1441 atgtgaaatagctgtgttcaataatgagatgtgttcagacatatacttccggtctgcgc 1500
 Oy 1420 agcaagatatacgaacacacgcgcacatatacttcttcttcttcttcttcttcttcttct 1479
 Db 1501 agcaagatatacgaacacacgcgcacatatacttcttcttcttcttcttcttcttcttct 1560
 Oy 1480 gagctatgacgcagcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 1539
 Db 1561 gagctatgacgcagcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 1620
 Oy 1540 ttccaggtgcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 1599
 Db 1621 ttccaggtgcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 1680
 Oy 1600 gggagacatgcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 1659
 Db 1681 gggagacatgcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 1740
 Oy 1660 tattaacagcagcctgagatgagacgagaaagatcgatgacatgagacatgagacatgagacat 1719
 Db 1741 tattaacagcagcctgagatgagacgagaaagatcgatgacatgagacatgagacatgagacat 1800
 Oy 1720 gacttcttcaaacatcagatgagcctggtgtgagatcctggtgtgtgtgtgtgtgtgtgtgt 1779
 Db 1801 gacttcttcaaacatcagatgagcctggtgtgagatcctggtgtgtgtgtgtgtgtgtgtgt 1860
 Oy 1780 gctaaagacaaatagtgatataacatgagaaagaaagaaatgagatttcttga 1839
 Db 1861 gctaaagacaaatagtgatataacatgagaaagaaagaaatgagatttcttga 1920
 Oy 1840 tctgcaaaatcttctgtagagagacgagaaacgttggatttctgtagtgtgtagagc 1899
 Db 1921 tctgcaaaatcttctgtagagagacgagaaacgttggatttctgtagtgtgtagagc 1980
 Oy 1900 ggcgc 1904
 Db 1981 ggcgc 1985

RESULT 7
 AAA48939 standard; cDNA: 1942 BP.
 ID AAA48939:
 AC AAA48939:
 XX 06-DEC-2000 (first entry)
 DT XX
 DE XX
 KW Soybean diacylglycerol acyltransferase cDNA #1.
 KW Diacylglycerol acyltransferase; corn, rice; soybean; wheat;
 KW triacylglycerol; herbicide; EC2.3.1.20; ss.
 XX Glycine max.
 OS XX
 XX Location/Qualifiers
 FH Key 29..1543
 FT CDS /tag= a
 FT /product= Diacylglycerol_acyltransferase
 XX
 XX WO200032756-A2.
 PN 08-JUN-2000.
 PD 01-DEC-1999: 99WO-US28354.
 PF 02-MAR-1998: 98US-0110602.
 PR 31-MAR-1999: 99US-0127111.
 XX (DUPO) DU POINT DE NEMOURS & CO E. I.
 PA
 XX

XX 22-JUN-2000.
PD
XX 16-DEC-1999; 99WO-CA01202.
PF
XX 17-DEC-1998; 98US-0112812.
PR
XX (CANNA) NAT RES COUNCIL CANADA.
PA
XX
PI Zou J, Taylor DC, Wei Y, Jako CC;
XX
XX MPI: 2000-431592/37.
DR P-PsDB: AAY96853.
XX
XX
PT New DNA encoding diacylglycerol acyltransferase from Arabidopsis
PT thaliana for transforming plants and regulating seed oil content, fatty
PT acid synthesis and seed oil acyl composition in commercial and crop
PT plants
XX
PS Claim 2: Page 67-68; 91pp; English.
XX
XX This DNA encodes Arabidopsis thaliana diacylglycerol acyltransferase
CC (DGAT). The DGAT and an insertion mutant (AS11) are useful for regulating
CC seed oil content, the ratio of diacylglycerol/triacylglycerol proportions
CC in seed oil, fatty acid synthesis, seed oil acyl composition, seed
CC size/weight and carbon flux into other seed components in commercial and
CC crop plants. The natural formation of triacylglycerols can be modified to
CC increase the yield in commercial plant oils or modify their composition
CC to achieve specific commercial improvements of plants and plant products.
XX
XX Sequence 5193 BP; 1582 A; 894 C; 1814 T; 0 other;

```

RESULT 11
ID AAA51485
AC AAA51485 standard; DNA; 5339 BP.
XX
XX AAA51485;
DT 09-OCT-2000 (first entry)
XX
DE A. thaliana AS11 diacylglycerol acyltransferase genomic DNA.
XX
XX DGAT; diacylglycerol acyltransferase; seed oil; fatty acid synthesis.
XX
XX size; weight; carbon flux; TAG1; insertion mutant; ds.
XX
XX Arabidopsis thaliana.
XX
FH Key
FH CDS
FT Location/Qualifiers
FT 1761..4926
FT /tag= a
FT 1761..2150
FT /tag= b
FT /partial
FT 2151..2234
FT /tag= c
FT 2235..2315
FT /tag= e
FT 2316..2380
FT /tag= f
FT 2370..2515
FT /tag= d
FT /note= "insertion region"
FT 2381..2461
FT /tag= g
FT /note= "duplication of exon 2"
FT 2462..2552
FT /tag= h
FT 2553..2622
FT /tag= i
FT 2623..2728
FT /tag= j
FT 2729..2804
FT /tag= k
FT 2805..2901
FT /tag= l
FT 2902..2970
FT /tag= m
FT 2971..3050
FT /tag= n
FT 3051..3174
FT /tag= o
FT 3175..3234
FT /tag= p
FT 3235..3332
FT /tag= q
FT 3333..3410
FT /tag= r
FT 3411..3512
FT /tag= s
FT 3513..3612
FT /tag= t
FT 3613..3749
FT /tag= u
FT 3750..3893
FT /tag= v
FT 3894..3975
FT /tag= w
FT 3976..4079
FT /tag= x
FT 4080..4100
FT /tag= y
FT 4101..4184
FT /tag= z
FT 4185..4247
FT /tag= aa

```

```

FT intron
FT /tag= ab
FT 4248..4330
FT /tag= ab
FT 4331..4378
FT /tag= ac
FT 4379..4501
FT /tag= ad
FT 4502..4564
FT /tag= ae
FT 4565..4645
FT /tag= af
FT 4646..4696
FT /tag= ag
FT 4697..4821
FT /tag= ah
FT 4822..4926
FT /tag= ai
FT /partial

```

WO200036114-A1.

22-JUN-2000.

16-DEC-1999; 99WO-CA01202.

17-DEC-1998; 98US-0112812.

(CANADA) NAT RES COUNCIL CANADA.

Zou J, Taylor DC, Wei Y, Juko CC;

WPI: 2000-431592/37.

P-PSDB; AAY96854.

New DNA encoding diacylglycerol acyltransferase from Arabidopsis thaliana for transforming plants and regulating seed oil content, fatty acid synthesis and seed oil acyl composition in commercial and crop plants

Disclosure: Page 79-81; 91pp; English.

This DNA encodes Arabidopsis thaliana ecotype Columbia mutant AS11 diacylglycerol acyltransferase (DGAT). The AS11 mutant TAG1 allele has a 147 bp insertion located at the central region of intron 2. The insertion is a duplication of a segment that is composed of 12 bp from the 3' end of intron 1, the entire sequence of exon 2 (81 bp) and 54 bp from the 5' end of intron 2. The DGAT and the insertion mutant (AS11) are useful for regulating seed oil content, the ratio of diacylglycerol to triacylglycerol proportions in seed oil, fatty acid synthesis, seed oil acyl composition, seed size/weight and carbon flux into other seed components in commercial and crop plants. The natural formation of triacylglycerols can be modified to increase the yield in commercial crop plants or modify their composition to achieve specific commercial improvements of plants and plant products.

Sequence 5339 BP; 1620 A; 919 C; 930 G; 1870 T; 0 other;

Query Match

Best Local Similarity 27.7%; Score 528; DB 21; Length 5339; Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 attcttagctctcctcaatcgcgtctctcctcattagatctgttctctt 60
DB 1623 attcttagctctcctcctcaatcgcgtctctcctcattagatctgttctctt 60
OY 61 tcaattctctgcatgctctctgattctctctgaagcgtctcttccgagcgtgtt 120
DB 1683 tcaattctctgcatgctctctgattctctctgaagcgtctcttccgagcgtgtt 120
OY 121 cgtcaaacgcttttgaatgagcatttggattcgtcggcggttaactagtgagagag 180
DB 1743 cgtcaaacgcttttgaatgagcatttggattcgtcggcggttaactagtgagagag 180

```


Query Match	19.98;	Score 379.4;	DB 21;	Length 1559;
Best Local Similarity	66.08;	Pred. No.	2.9e-100;	
Matches 576; Conservative	0.			

b
1 ttttggttaatgcatcattcgagactgcacgcgtctatgttgtccctagtaccttcta 60
tcgatcttcccttttatcttgcctctctctctctctctctctctctctctctctctct

754 gaacgtgtgcatcttcttcatatattatcacatgacagagtttctgtatcaatt 813
121 tttctcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt

181 ctgcgtatctcttaagtgtgattcgtcagatttcatcagagcttctatattatattat

241 tgcattgttgcgtgaagcttgcatacacaatacattatataagaacatg 333
934 gccaatgcagctgataaggccaatcctgaagctcctactacattatg 300

987 -----gagcttgcgatattcatgtcgcctccacaatgtgttatcagca 360
||| | |||| ||||||
361 gctccactctctgggaqtcttaacatactttttt 1032

421 agtattctctcgacaccctatgttagaaaaggtgctgcgtcaagttatctctac 1092
1093 gtcataatcaccgcaattca

[illegible]

601 tcaatccaagaatgctacactgctgctgcaatgcttattgacctttacacata

601 aacataactgcgcgagatctcttcgatattgttgtaaccgagaattctacaagaactgttgyaat 720

393 cgacatatataacttccgctgtcgcagcaagataccacaagaactcgcatattcatt 780
|| ||||||| | | |||||
781 cgtatatatatatttcctttcttatattt | ||| |||||

841 | ||| | | gggccccccagagcta 1485
tcgtctcttgcttctgcgtgactcactcaggta 873

PA (CANADA) NAT RES COUNCIL CANADA.

XX Zou J, Taylor DC, Wei Y, Jako CC;

PI MPI: 2000-431592/37.

DR MPI: 2000-431592/37.

PT New DNA encoding diacylglycerol acyltransferase from Arabidopsis
PT thaliana for transforming plants and regulating seed oil content, fatty
PT acid synthesis and seed oil acyl composition in commercial and crop
XX plants

PS Disclosure: Page 72; 91pp; English.

CC This Arabidopsis thaliana EST (expressed sequence tag) has a deduced
CC amino acid sequence showing 41 percent identity to that of the yeast
CC acyl-CoA cholesterol acyltransferase. The cDNA was used to isolate a
CC clone encoding A. thaliana diacylglycerol acyltransferase (DGAT). The
CC DGAT and an insertion mutant (AS11) are useful for regulating seed oil
CC content. The ratio of diacylglycerol/triacylglycerol proportions in seed
CC oil, fatty acid synthesis, seed oil acyl composition, seed size/weight
CC and carbon flux into other seed components in commercial and crop plants.
CC The natural formation of triacylglycerols can be modified to increase
CC the yield in commercial plant oils or modify their composition to achieve
CC specific commercial improvements of plants and plant products.

CC Sequence 629 BP; 150 A; 142 C; 164 G; 157 T; 16 other;

SO

Query Match

Best Local Similarity 17.7%; Score 336.4; DB 21; Length 629;

Matches 381; Conservative 94.8%; Pred. No. 6.6e-88;

Mismatches 381; Conservative 0; Mismatches 16; Indels 5; Gaps 3;

OY 1047 tgcattatagcgaagggttggtggtgcgtcaatttgcaaacgtgcatattaccgg 1106
DB 1 tgcattatagcgaagggttggtggtgcgtcaatttgcaaacgtgcatattaccgg 60
OY 1107 attcattggaattataatagaacaataataatcatttgcaggaactcaagaatcc 1166
DB 61 attcattggaattataatagaacaataataatcatttgcaggaactcaagaatcc 120
OY 1167 ttgaaaggcgatctctatataatgctatgaaagggttgaaagcttcaattccaattt 1226
DB 121 ttgaaaggcgatctctatataatgctatgaaagggttgaaagcttcaattccaattt 180
OY 1227 atattgtggtcgtcgtatctctatgcttctccaccttggttaacatattgcaga 1286
DB 181 atattgtggtcgtcgtatctctatgcttctccaccttggttaacatattgcaga 240
OY 1287 gcttcctgctcggtggtatgcttcaacaagaattggtggaatgcaaaaagtgtgg 1346
DB 241 gcttcctgctcggtggtatgcttcaacaagaattggtggaatgcaaaaagtgtgg 300
OY 1347 agattact-ggaagaattggaatgcttctcaataatg--atggttcacatata 1403
DB 301 agattactggaagaattggaatgcttctcaataatg--atggttcacatata 360
OY 1404 C--ttccgtggttcgcaagaatgcaagaactcgtcc 1443
DB 361 ccttcctggttcgcaagaatgcaagaactcgtcc 402

Search completed: May 13, 2002, 07:32:50
Job time: 14454 sec

Mon May 13 10:17:53 2002

us-09-623-514a-3.rst

Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 04:58:46 ; Search time 5105.53 Seconds
(without alignments)
13728.190 Million cell updates/sec

Title: US-09-623-514A-3
Perfect score: 5193
Sequence: 1 gtcacgacccattcttccc.....tggtgtcccttgggga 5193

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: EST:

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estum:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_hic:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pin:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	661.4	12.7	663	12	AO958056 LERA062TF
2	582.6	11.2	605	12	AO958057 LERA062TR
3	537.8	10.4	541	12	CNS000WZ
4	433	8.3	746	12	BH466416 BOHEK18TR
5	374.4	7.2	399	10	BE523720 M40H1STM
6	276.6	5.3	545	9	AI992543
7	211.4	4.1	852	12	BH464564
8	177.4	3.4	823	12	BH460807
9	163.8	3.2	525	12	BH546158 BOGTW37TR
10	141.8	2.7	629	9	AA042298 24635 CD4
11	131.4	2.5	1101	12	CNS00KAE
12	124	2.4	661	12	BH540136 BOHRT87TR
13	121.4	2.3	829	12	CNS03LDB
14	119.8	2.3	1101	12	CNS00EVL
15	118	2.3	261	10	BE525637
16	118	2.3	660	12	BH183498
17	118	2.3	660	12	CNS07ONT

18	118	2.3	862	10	BG321213
19	117.8	2.3	982	12	AQ325799
20	117.6	2.3	614	12	AL104915 Drosophila
21	116.8	2.2	1101	12	CNS00EVL
22	115.6	2.2	1101	12	CNS0005X
23	115.6	2.2	1101	12	CNS017V2
24	115.2	2.2	879	12	CNS01JRG
25	114.8	2.2	737	12	CNS008HU
26	114	2.2	1029	12	CNS01ZGM
27	112.6	2.2	886	12	BH177277
28	112.6	2.2	886	12	CNS07JDX
29	112.6	2.2	945	12	CNS04DOK
30	111.6	2.1	1101	12	CNS0039G
31	110.8	2.1	260	10	BE525720
32	110.8	2.1	990	12	CNS006OT
33	110	2.1	580	12	CNS008HN
34	110	2.1	1036	12	CNS00599
35	110	2.1	1139	12	AO897537
36	109.8	2.1	1135	12	CNS033GQ
37	109.8	2.1	1201	12	CNS0167M
38	109.6	2.1	1101	12	CNS0039G
39	109.2	2.1	1101	12	CNS003DX
40	108.8	2.1	1885	10	BE420745
41	108.6	2.1	1079	12	CNS0351X
42	108	2.1	1159	12	CNS015XR
43	107.2	2.1	1101	12	CNS0021J
44	105.4	2.0	922	12	AZ548363
45	105.8	2.0	907	12	CNS021J4

ALIGNMENTS

RESULT 1
LOCUS AO958056 663 bp DNA linear GSS 28-JAN-2000
DEFINITION LERA Arabidopsis thaliana genomic clone LERA062, DNA sequence.
ACCESSION AO958056.1 GI:6785757
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Unpublished (2000)
Contact: Xiaoying Lin
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: ateligr.org
For additional information, see <http://www.tigr.org/tdb/at.html>
Seq primer: TF
Class: shotgun.
FEATURES
source
1..663
/organism="Arabidopsis thaliana"
/strain="Landsberg erecta"
/db_xref="taxon:3702"
/clone="LERA062"
/clone="LERA"
/note="Organ: Leaf; Vector: pROSL; Total genomic DNA was sheared to 0.9-1 kbp before ligation."
BASE COUNT
191 a 129 c 123 g 220 t
ORIGIN

Query Match 12.7%: Score 661.4; DB 12; Length 663;
 Best Local Similarity 99.8%; Pred. No. 4.1e-12;
 Matches 662; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4232 ggtatcaacttcttctacattgcccgtctctcaacgacatgaacacgctagtgaaag 4291
 Db 1 ggtatcaacttcttctacattgcccgtctctcaacgacatgaacacgctagtgaaag 4291
 QY 4292 aatgcaatattcaaaagcattgttttcttacttaacagatcttggtacaaattcccttg 4351
 Db 61 aatgcaatattcaaaagcattgttttcttacttaacagatcttggtacaaattcccttg 4351
 QY 4352 aacgcttgatcgacgtctctctgctctctcaacgacatgaacacgctagtgaaag 4411
 Db 121 aacgcttgatcgacgtctctctgctctctcaacgacatgaacacgctagtgaaag 4411
 QY 4412 gttcaggttaaaaaaatacttaacacgtctgacgtcttcttacttaacatctcat 4471
 Db 181 gttcaggttaaaaaaatacttaacacgtctgacgtcttcttacttaacatctcat 4471
 QY 4472 attcgaacacacacattgtttgaagtaggtgcttctgctctcaacatctcat 4531
 Db 241 attcgaacacacacattgtttgaagtaggtgcttctgctctcaacatctcat 4531
 QY 4532 ggaagagtttgctcaacgctatgctctcaacacacacacacacacacacacacacac 4591
 Db 301 ggaagagtttgctcaacgctatgctctcaacacacacacacacacacacacacacac 4591
 QY 4592 tcttcatagcctgaacacatttaacacacacacacacacacacacacacacacacacac 4651
 Db 361 tcttcatagcctgaacacatttaacacacacacacacacacacacacacacacacacac 4651
 QY 4652 ttggaatggaatcatalattatagtggtggaacacacacacacacacacacacacac 4711
 Db 421 ttggaatggaatcatalattatagtggtggaacacacacacacacacacacacacac 4711
 QY 4712 ttcgaac 480
 Db 481 ttcgaac 480
 QY 4772 atgcatgaac 4831
 Db 541 atgcatgaac 4831
 QY 4832 ctccgctgattgttggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 4891
 Db 601 ctccgctgattgttggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 4891
 QY 4892 gaa 4894
 Db 661 gaa 663

RESULT 2
 AO958057/c 605 bp DNA Linear GSS 28-JAN-2000
 DEFINITION LERA62R LERA Arabidopsis thaliana genomic clone LERA62, DNA
 LOCUS sequence.
 ACCESSION AO958057
 VERSION AO958057.1 GI:6785758
 KEYWORDS GSS.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 605)
 AUTHORS Buell, C.R., Lin, X., Pai, G., Barnstead, M., Bowman, C., Utebach, T.,
 Feldblyum, T., Liang, F., Creasy, T., and Fraser, C. M.
 TITLE Genomic survey sequencing of Landsberg erecta ecotype of
 Arabidopsis thaliana and identification of sequence-based

JOURNAL polymorphisms
 COMMENT Unpublished (2000)
 Contact: Xiaoying Lin
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: atetigr.org
 For additional information, see <http://www.tigr.org/tdb/at/at.html>
 Seq primer: TR
 Class: Shotgun.

FEATURES
 source
 location/Qualifiers
 1. 605
 /organism="Arabidopsis thaliana"
 /strain="Landsberg erecta"
 /db_xref="taxon:3702"
 /clone="LERA62"
 /note="Organ: Leaf; Vector: PHOS1; Total genomic DNA was
 sheared to 0.9-1 kbp before ligation."

BASE COUNT 188 a 122 c 90 g 205 t
 ORIGIN

Query Match 11.2%: Score 582.6; DB 12; Length 605;
 Best Local Similarity 99.2%; Pred. No. 1.9e-62;
 Matches 596; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 4526 tctacagaaagtttgctcaa-cggtatgctctcaaacacacacacacacacacacacacac 4584
 Db 602 tctacagaaagtttgctcaa-cggtatgctctcaaacacacacacacacacacacacacac 4584
 QY 4585 aactcttcttcaagcctgaac 4644
 Db 542 aactcttcttcaagcctgaac 4644
 QY 4645 gatctgtttggaatggaatcatalattagtggtggaacacacacacacacacacacacacac 4704
 Db 482 gatctgtttggaatggaatcatalattagtggtggaacacacacacacacacacacacacac 4704
 QY 4705 ctgcatcttggaac 4764
 Db 422 ctgcatcttggaac 4764
 QY 4765 aggatcgatgcatgaac 4824
 Db 362 aggatcgatgcatgaac 4824
 QY 4825 gtgagatccggttgatggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 4884
 Db 302 gtgagatccggttgatggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 4884
 QY 4885 catlgaagaagaagaagaatagagttgtatctgcaacacacacacacacacacacacacacac 4944
 Db 242 catlgaagaagaagaagaatagagttgtatctgcaacacacacacacacacacacacacacac 4944
 QY 4945 cgaacccgttgatgttttgaatggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 5004
 Db 182 cgaacccgttgatgttttgaatggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 5004
 QY 5005 attgttaccac 5064
 Db 122 attgttaccac 5064
 QY 5065 ttctctgtttgac 5124
 Db 62 ttctctgtttgac 5124
 QY 5125 a 5125
 Db 2 A 2

Db	454	CTTACACTTGGAAACGGTTT	TTTTGTTAAACCTCGCATCTATCTTTT	-CCACGCCAAGCT	396
OY	3272	atcaagttctgcatgatalacaggaaggtctgggtggcctgcgtcaatattgcaaaactgtgca	3331		
Db	395	ATCCACGCTCCCATCATGATACGGAAGGTTGGGTGGCTGTCATATTGGCAAACTGGTCA	3336		
OY	3332	tattcaacggattcatcagggaattataatagaacaagtcgtcttccaacttgcattat	3391		
Db	335	TATTCCACTGGACATCAAGGATTTATAAATAGACAGGTCGGTTTCAACATCTTCTTTT	276		
OY	3392	agttt--tctctgtgtaaaatcatcatccctgcgtgtgc-accaactgtacattcatcttc	3448		
Db	275	ATTATATAATCCCTGGGAAGATCACCATTTCTGCAATTCGACCGCGTTGGCTTATCTTTCT	216		
OY	3449	tttgttgaat-ttggcagtaataaacatccattgttcaggaactcaagaatccattgaa	3507		
Db	215	TTTGTACTCTCTTGGCAGTATTAATTCATTTGTAAGSAATCTAAACATCCGTGAA	156		
OY	3508	aggcgatctctatatatgcatattgaagaagtgctgaagcttcaagttccaatttatagt	3567		
Db	155	AGGGATCTTTTATAACCTATTGAAGAAGCTTGAAGCTTTAGTCCAAATCTATAAGT	96		
OY	3568	gtgcctctgcattgtctcaactgtcttccaaccttg----gtatgctgtgataccatctc	3623		
Db	95	GTCGCTCTGCATGTTCTACTACGCTTTCTTCCACCTTTGGTATGATGCCGTGATCCCTTCTC	36		
OY	3624	tttaaaataatttgcnaattctgnaaaacaggaanaag	3660		
Db	35	TCTCAACATAAAT--CCAAAGAGAAACAAACAGAAAAG	1		

RESULT	5
LOCUS	BE523720
DEFINITION	BE523720 399 bp mRNA linear EST 19-MAR-2001
ACCESSION	M40H1M Arabidopsis developing seed Arabidopsis thaliana cDNA
VERSION	BE523720
KEYWORDS	BE523720.1 GI:9781698
SOURCE	EST.
	tale cross

BASE COUNT	ORIGIN	
72 a	104 c	85 g 138 t

Query Match	7.2#;	Score 374.4;	DB 10;	Length 399;
Best Local Similarity	98.4#;	Pred. 0; 7.1e-37;		
Matches 378;	Conservative	0;	Mismatches 6;	Indels 0;
				Gaps
QY 1608	atccatcccccacgcgattcttgatgcctcttccttcaatccgcgtctcttcctctcatg	1667		
Db 16	ATCTCATTCCTCCACGAGATTCTTACGCTTCTTCCAAATCGCTCTTCCCTCCATTAAG	75		
QY 1668	atcttgctctctcttcaattctctctgatgctctcgatctctctctgaagcctcttt	1727		
Db 76	ATTCTGTTCTCTCTTTCAAATTCCTCTGCATGCTTCTGCATTCCTCTCGAGCCCTCTTTT	135		
QY 1728	ctcccgacgcgtgttctgcacaagctcttcgaaatggcgatcttgatctctgcgtg	1787		
Db 136	CTCCCGACGCTGTTTGTGTCAAACGCTTTCCAAATGGGAGATTGGATTCTGCTGCGGTT	195		
QY 1788	actcagcgtgacggagaagcgttgccggagaatcgtgcgactcttgatcttgatgcgttcgacgg	1847		
Db 196	ACTACGCTTACTGAAACAGGCTGGGAGAGATTCTGCAATCTTTAGGCTTCGTCACGG	255		
QY 1848	aaatcgagatcggaattctctcaacggaattctctctcgctgcgataataattcct	1907		
Db 256	AAATCGAATCGGATTCTTCTTAACGAGCTTCTCTCTGCTTCGATATTAATTTCTCT	315		
QY 1908	tccgattgcttgagctcccccgcgaagttaggatcggatctgattcgcgtcttcaagat	1967		
Db 316	TCCGATTGATGTTGGAGCTCCCGCGACGTTAGGATCTGATTGATTTCTGTTAAAGAT	375		
QY 1968	gacgcctcggagaagacgaatttg	1991		
Db 376	GACGCTCACGGAACGCGCAATTGG	399		

RESULT	6	
LOCUS	A1992543/C	
DEFINITION	A1992543	545 bp mRNA
ACCESSION	70158181 A.	thaliana, Ohio State clone set Arabidopsis thaliana
VERSION	GDNA clone	70158181, mRNA sequence.
KEYWORDS	A1992543.1	GI:5839448
SOURCE	EST.	
ORGANISM	thale cress.	
	Arabidopsis thaliana	
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;	
	Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.	
REFERENCE	1 (bases 1 to 545)	
AUTHORS	Chen,J., Momiyama,M., Chan,E., Mooney,M., Carroon,B., Gilliland,D., Wang,X., Hillman,J., Guebler,K., Kim,C., Doyle,M., Brzoska,P., Gorone,G., Burns,D., Griffin,J., Mouanoutoua,M., Nguyen,D., Tan,R., Rose,M., Warren,B., Ton,B., Kastury,K., Borillo,C., Carpio,T., Policky,J., Suzuki,G., Argentine,C., Shah,S., Noblitz,A., Murty,L., Turner,C., Kikorian,S., Elder,L. and Hanson,D.	
TITLE	Arabidopsis thaliana Gene Expression Microarray	
COMMENT	Unpublished (1999)	

BASE COUNT 146 a 125 c 120 g 154 t
 ORIGIN /clone="701558181"
 /clone.lib="A. thaliana, Ohio State clone set"
 /note="cDNA library was made from selected clones from the
 Arabidopsis thaliana Ohio State clone set."

Query Match 5.3%; Score 276.6; DB 9; Length 545;
 Best Local Similarity 95.3%; Pred. No. 5.9e-25;
 Matches 285; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 4675 ggtggggaatgatgatcttcgtgctacatcttcgcaatcttcgcaacacgagtggtgtc 4734
 ||||||| ||||||| ||| ||| ||||||| ||||||| ||||||| ||||||| |||||||
 DB 299 GGTGGGAACATGATCTCCGTTCAATCTCTCGCATTTTCGGACACCCGATGTGTGCT 240
 QY 4735 tcttattaccagacttgatgaacggaagagtgatgcatgaacacgagtgctcaa 4794
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 DB 239 CCTTATTACGACGACCTGATGACCGAAGATCATGTCATGAACAACACTTTCAA 180
 QY 4795 aaatgacttctcaacatcatatgagcctcgttgatctcctgtgagtggtgtc 4854
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 DB 179 AAATGACTTTTTCACACATGATGCGCTGTCGATCCCGTTGATGTTGTGTGCTTC 120
 QY 4855 tgatgctaaacgcaaatagtgctataaccattggaagaaagaaatagagtgct 4914
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 DB 119 TGATGCTAAACGACAAATAGTGTATTAACCATGGAAGAAAGAAATTAAGATTGT 60
 QY 4915 tgtatctcaaaaatttggtagagacacgcaacccgttgatgttggatgtgta 4973
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 DB 59 TGATCTCAAAAATTTTGGTAGAGACGCAACCCGTTGATGATTTGTATGTGTGTA 1

RESULT 7
 BH464564/c 852 bp DNA linear GSS 13-DEC-2001
 LOCUS BOGZG21TF BOGZ Brassica oleracea genomic clone BOGZ21, DNA
 DEFINITION
 sequence.
 ACCESSION BH464564
 VERSION BH464564.1 GI:17660969
 KEYWORDS GSS.
 SOURCE Brassica oleracea.
 ORGANISM Brassica oleracea.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 852)
 REFERENCE Town,C.D., Van Aken,S., Uterback,T. and Fraser,C.M.
 Whole genome shotgun sequencing of Brassica oleracea
 unpublished (2001)
 JOURNAL Other_GSSs: BOGZG21TF
 COMMENT Contact: Chris Town

TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TF
 Class: Sheared ends.

FEATURES
 source
 1. 852
 /organism="Brassica oleracea"
 /strain="TO1000DH3"
 /db_xref="taxon:3712"
 /clone="BOGZG21"
 /clone.lib="BOGZ"
 /note="Vector: PHOS1, Site_1. BstXI, 2-3 kb sheared
 genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 318 a 152 c 162 g 220 t
 ORIGIN

Query Match 4.1%; Score 211.4; DB 12; Length 852;
 Best Local Similarity 67.0%; Pred. No. 4.7e-17;
 Matches 428; Conservative 0; Mismatches 161; Indels 50; Gaps 7;

QY 2561 tgatcttctctctgtgtgacatatacccttgatctcttcttgctgcttcagc 2620
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 DB 756 TGCCTCTTTTATGTTGCTGCTCTCTCTATCAATAGTCCCTTTGGC--CCTTTACCG 699
 QY 2621 ttgagaatctgttacttgaataatatacagaacccgtgtgaatactatctccag 2680
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 DB 698 TCGAGAAATGATCTTCAAGAAATGCTCTCTCTGACCTGATCTGTAGTAACTA 639
 QY 2681 ccattactgaatttlatatgaagaagtttgatcatalgaagaactcaagttctgt 2740
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 DB 638 CCATTACTGATTTGTACTAGAA-----ACTTATTACTTTCTTT 597
 QY 2741 ttgaatctcaaggtgttcattcttcttcaatt-attaccacatgacagagttc 2799
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 DB 596 CTGATACTACTGAAGGTGTGATCATCTTTCAATTTGTTATCACCATGACAGTCTT 537
 QY 2800 gtaaccagttacgtacaccataagtgatactgtttcttggtctcagttgtgatactg 2859
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 DB 536 GTATCCAGT---CTATATTAAGCTATGTGTGTGTTCTGGTGTA-----TACTCTT 485
 QY 2860 ttcttaagttagttgctgaacccgtgacatctgaaatggaaggtgtgattgtctt 2919
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 484 TTTTATTTAGTTAGTTGCTAACCTTGCGATCTTGAATGAGAGTGTGATTCGCCAT 425
 QY 2920 ttatacaggtgacacttgatgtctctc-----actgacatgtgtgctaaagt 2971
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 DB 424 TCTGTCAGTTGTACATTTGATGCTCTCGAATCAGTACTGATGTTGTTGATGAGTTG 365
 QY 2972 gttcttattgctatactagatctatgacataagatccctagccaatgacgtgataagta 3031
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 DB 364 GTTCTTAGGCTACTACTACTACACTAACCTTAAGAACTCTAC-----GTACATT 318
 QY 3032 aaatacgaanaagaagcgtatgattagtcactgacgtgtgtactgtttaaacaac 3091
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 317 AGTTACTGTGCTATGTTACTTATTTGACAACTTGCAATTGTGTTAAACCGAC 258
 QY 3092 actgtatgaactttagccaactcctgaagctctcactaactgaactgtgaagactgtg 3151
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 257 AGTGTGAACACTTTAGGCTAATCTGAAGTCTCTACTGATGCTTGAAGAACTGTG 198
 QY 3152 cataattcaagtgcctcccaactgtgtatcaggtaa 3190
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 DB 197 CATATTTCATGTGCTCTCTTACATTGTTATCTAGTAA 159

RESULT 8
 BH460807/c 823 bp DNA linear GSS 13-DEC-2001
 LOCUS BOGOV68TF BOGO Brassica oleracea genomic clone BOGOV68, DNA
 DEFINITION
 sequence.
 ACCESSION BH460807
 VERSION BH460807.1 GI:17650552
 KEYWORDS GSS.
 SOURCE Brassica oleracea.
 ORGANISM Brassica oleracea.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 823)
 REFERENCE Town,C.D., Van Aken,S., Uterback,T. and Fraser,C.M.
 Whole genome shotgun sequencing of Brassica oleracea
 unpublished (2001)
 JOURNAL Other_GSSs: BOGOV68TF
 COMMENT Contact: Chris Town

TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org

COLEMAN, J. W. 1964. The ecology of the Colorado Desert, California. *Ecology* 45: 1-12.

'E.' Onitoge, O., Kalknel, N., S
, E. and Somerville, C.

102(b)

TITLE Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones
JOURNAL Plant Physiol. 106, 1241-1255 (1994)
MEDLINE 93148729
COMMENT On Sep 19, 1997 this sequence version replaced g1:1520456.
Contact: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E. Lansing, MI
Tel: 517-353-0854
Fax: 517-353-9168
Email: 22313tcn@lhm.cl.msu.edu
Seq primer: 77.

FEATURES
source
1. 629
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="E6B27"
/tissue="seedling hypocotyl"
/dev_stage="3 day-old"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: EcoRI; Using 5 ug of polyadenylated mRNA from 3 day-old Arabidopsis thaliana (Columbia) seedling hypocotyls as template and oligo d(T) as primer, first strand synthesis was catalyzed by Moloney murine leukemia virus reverse transcriptase (Pharmacia). Second-strand cDNA was made using the procedure of Gubler and Hoffman (1983) except that DNA ligase was omitted. After the second strand reaction, the ends of the cDNA were made blunt with Klenow fragment and EcoRI/MotI adapters (Pharmacia) were ligated to each end. The cDNA was purified from unligated adapters by spin-column chromatography using sephacryl s-300 and size-fractionated on a 1% low melting point mini-gel. Size selected cDNAs (0.5 - 1 kb) were removed from the gel using agarase (New England Biolabs), phenol:chloroform extracted and precipitated using 0.3 M NaOAc (pH 7)/ethanol. A portion of each cDNA size-fraction (0.1 ug) was co-precipitated with 1 ug of lambda-zapit (Stratagene/EcoRI digested, dephosphorylated arms and then ligated in a volume of 4 ul overnight. Each ligation mix was packaged in vitro using Gigapack II gold packaging extract (Stratagene). We have determined that although first strand cDNA synthesis was initiated using a tail, almost all of the cDNAs begin 8-10 bp from the poly-A tail. The reason for the loss of the poly-A tail is most likely due to lower than anticipated nucleotide levels during the Klenow repair of ragged ends before the addition of linkers (3'-5' exo instead of 5'-3' pol). When this library is used please reference the ABC and: Kieber, J. et al. (1993) Cell 72:427-441."

BASE COUNT 150 a 142 c 164 g 157 t 16 others
ORIGIN

Query Match 2.7%: Score 141.8; DB 9; Length 629;
Best Local Similarity 95.4%; Pred. No. 1.8e-08;
Matches 146; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3453 ttaactttgagcagatataaataatccatctgcaggaactcaagaacatcccttcggaagcgc 3512
DB 72 TTATATATGAGCAATATATATATATCTATGTCAGAGCACTCAAGCATCCTTANAAAGGCG 131
QY 3513 atctctatagcattgaagaagtggtgaagcttcagttccaaattataltggtg 3572
DB 132 ATCTCTATATGCTATTTAAAGAGTGTGGAAGCTTTCAGTTCCTCAATATATATGTCGCG 191
QY 3573 tctgcagtgctactgctctccaccttggt 3605
DB 192 TCTGCATGTTTACTGCTTCTCCACCTTGT 224

RESULT 11
CNS00KAE 1101 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC:
DEFINITION BACR3905 of Rpci-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL077628.1 GI:4956903
VERSION GSS.
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
REFERENCE
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoler in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named Rpci-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the library p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source
1. 1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR3905"
/note="end : TET3"

BASE COUNT 374 a 59 c 46 g 399 t 223 others
ORIGIN

Query Match 2.5%: Score 131.4; DB 12; Length 1101;
Best Local Similarity 39.8%; Pred. No. 2.7e-07;
Matches 384; Conservative 80; Mismatches 481; Indels 19; Gaps 3;

QY 809 taaatcatcattacatcttgatccaaacaatcattcatatctatttgatggttaag 868
DB 2 TATWMMNTMMWMTTATATATTTTATTTATTTATATATATATATATATATATAT 61
QY 869 aaacactattacaggttacaataatataaagaatggttgtaaaaaaagataca 928
DB 62 NANTATTAANTNTNTANANTNMNTATATATAAANCTNCTATATAGATATATATAT 121
QY 929 agtgaattctcttgccaataataaattgactcttcaatataatgacttttgga 988
DB 122 NNTNTNATATATATNTNTNANANANNTTATATATATANTANTANANANANIN 181
QY 989 catgattacagaataatcatcatcacaacttccaagttata-----ataaa 1037
DB 182 ANATATATTAANANTNTNTANTNTATAGGANNMNAATATANTATANTNTNTN 241
QY 1038 tacattccaagaactagttctctcttaaaatattcttaaaagtgaacaaagataca 1097
DB 242 TNAATATATATATANANANANATATANANANATANTANANTATAGGAAATTTAT 301
QY 1098 cataaatcagaaaaagtagaagttgactcttcttgccaataaataattgactaaa 1157

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 07:20:45 ; Search time 150.64 seconds
(without alignments)
8467.703 Million cell updates/sec

Title: US-09-623-514A-3
Perfect score: 5193
Sequence: 1 gctcagcaccatcttcc.....tggtgtcccggtttgggga 5193

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 38353 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued_Patents_NA:*
2: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	141.8	2.7	629	4	US-09-103-754A-3
2	100.6	1.9	19124	2	US-08-487-826B-13
3	87.8	1.7	19124	2	US-08-487-826B-13
4	83.6	1.6	7218	1	US-08-232-463-14
5	82	1.6	6124	4	US-08-213-419B-3
6	81.2	1.6	1511	1	US-07-991-867B-8
7	81.2	1.6	1511	1	US-08-107-755A-8
8	81.2	1.6	1511	2	US-08-544-332-8
9	79.4	1.5	1511	1	US-07-991-867B-8
10	79.4	1.5	1511	1	US-08-107-755A-8
11	79.4	1.5	1511	2	US-08-544-332-8
12	79.4	1.5	240	1	US-08-628-417-6
13	75.4	1.5	6243	2	US-09-056-075-1
14	74	1.4	658	4	US-08-998-416-595
15	74	1.4	8920	2	US-08-446-855A-1
16	74	1.4	8920	2	US-09-150-741-1
17	73.8	1.4	837	4	US-08-998-416-288
18	71.8	1.4	615	4	US-08-998-416-186
19	71.6	1.4	636	4	US-08-998-416-1137
20	71.6	1.4	2447	2	US-09-014-969-14
21	70.4	1.4	4818	3	US-08-947-823-1
22	70.4	1.4	51952	3	US-08-852-629-11
23	70.2	1.4	4810	3	US-08-852-629-11
24	68.8	1.3	854	3	US-08-852-629-15
25	68	1.3	6124	4	US-08-998-416-534
26	67.6	1.3	3701	4	US-08-213-419B-3
27	67.6	1.3	3701	4	US-08-845-258-10

28	67.6	1.3	3701	4	US-08-990-571-10	Sequence 10, Appl
29	67.6	1.3	3701	4	US-08-723-142A-10	Sequence 10, Appl
30	66	1.3	665	2	US-08-883-795A-36	Sequence 36, Appl
31	65.6	1.3	860	4	US-08-998-416-287	Sequence 287, Appl
32	65.4	1.3	724	4	US-08-998-416-683	Sequence 683, Appl
33	65.4	1.3	5852	1	US-07-867-106-2	Sequence 2, Appl
34	65.2	1.3	615	4	US-08-998-416-186	Sequence 186, Appl
35	65	1.3	660	1	US-07-991-867B-32	Sequence 32, Appl
36	65	1.3	660	1	US-08-107-755A-32	Sequence 32, Appl
37	65	1.3	660	2	US-08-544-332-32	Sequence 32, Appl
38	65	1.3	854	4	US-08-998-416-534	Sequence 534, Appl
39	65	1.3	9048	3	US-08-973-273-4	Sequence 4, Appl
40	64.6	1.2	1431	4	US-09-316-083-2	Sequence 36, Appl
41	64.2	1.2	665	2	US-08-883-795A-36	Sequence 36, Appl
42	64.2	1.2	3095	6	5231168-1	Patent No. 5231168
43	64	1.2	5852	1	US-07-867-106-2	Sequence 2, Appl
44	64	1.2	6768	4	US-08-107-755A-1	Sequence 1, Appl
45	64	1.2	8457	1	US-07-991-867B-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-103-754A-3
Sequence 3, Application US/09103754A
Patent No. 6344548
GENERAL INFORMATION:
APPLICANT: Farese, Robert
APPLICANT: Cases, Sylvaine
APPLICANT: Smith, Steven
APPLICANT: Erickson, Sandra
TITLE OF INVENTION: Diacylglycerol O-acyltran
sferase
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: Bozicevic & Reed
STREET: 285 Hamilton Avenue, Suite 200
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/103,754A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Field, Bret E
REGISTRATION NUMBER: 37,620
REFERENCE/DOCKET NUMBER: 6510-105P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650 327 3400
TELEFAX: 650 327 3231
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 629 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-103-754A-3
Query Match 2.7%; Score 141.8; DB 4; Length 629;

Best Local Similarity 95.4%; Pred. No. 8.3e-19;
Matches 146; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 3453 ttacatttggcagatgaataactctatctgaggaactcaagacatcttgaagcg 3512
Db 72 TTATATATGAGCAATATATATATATATATATATATATATATATATATAT
OY 3513 attctctatgctatgaaagagtggtgaagcttcagttcgaattatggtgac 3572
Db 132 ATCTCTATATGCTATGTAAGAGCTTGAAGCTTTCAGTTCCAAATTTATGTGCG 191
OY 3573 tctgacgttctactgctctctcacttctggt 3605
Db 192 TCTGATGTTCTACTGCTCTCTCCACCTTTGGT 224

RESULT 2
US-08-487-826B-13

Sequence 13, Application US/08487826B
Patent No. 5993827
GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESS: Knobbe Martens Olson & Bear
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08-487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israel, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ. ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 19124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-487-826B-13

Query Match 1.9%; Score 100.6; DB 2; Length 19124;
Best Local Similarity 46.2%; Pred No. 1.4e-10;
Matches 595; Conservative 0; Mismatches 669; Indels 23; Gaps 7;

OY 183 tttactaagaataattttagtttctggttgaagcagatgattgtaacttaaa 242
Db 5832 TAAAAAATAGATACCAATATGATGATATATACGAGATATATTTGTATATAATATA 5891

OY 243 aaaaatgattgtaattcttaaaaatgtagaggtacacatcaactcagcagatc 302
Db 5892 TAT
OY 303 aaaaacgtaatacagcagatgtagaat--caatagaagatgtagatctttaa 360
Db 5952 ATCCCTTTAT
OY 361 atatcgatctcaaatgcataactcactagaagaatgtaattctttagtgctacataa 420
Db 6012 ATATATCAAAAT
OY 421 taataatataatataatataatataatctgtatataatgctgacaaaatgcca 480
Db 6072 TGCAT
OY 481 gtcaaaacccatgactgatacaactaaactaagtcgagatgtaactaactaagtcgag 540
Db 6131 TTTTAT
OY 541 agtataatctcattatgcttctatctacttcaaacacggaatagatattatagatc 600
Db 6187 AT
OY 601 caaaaagtagattggtatataatgaaagattggaattcaatcaatcaagatc 660
Db 6247 TAT
OY 661 aaagtaactccatlaataatcagtcgctgtagaaagcccaatgattgtagaattg 720
Db 6307 AATGCTAT
OY 721 gaaagttatlaataatgagatcttcttcttcttcttcttcttcttcttcttcttct 780
Db 6360 TAAAT
OY 781 ttcaaatatcaactatctactactactactactactactactactactactactact 840
Db 6420 AT
OY 841 catctatctactcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 6479
Db 6480 TGAATGCTAT
OY 901 aggat-tgltgtagaataaaagtagaagtagaagtagaagtagaagtagaagtaga 959
Db 6540 AT
OY 960 acttcttaataatgacttcttcttcttcttcttcttcttcttcttcttcttcttct 6599
Db 6600 AT
OY 1020 ttcaagttatataaat-acttcaagagactgagtccttcttcttcttcttcttct 1078
Db 6660 TGTCTCTTTGTTATCTCTATATATATATATATATATATATATATATATATATAT
OY 1079 aagtagaagtagaagtagaagtagaagtagaagtagaagtagaagtagaagtaga 6719
Db 6720 AAAT
OY 1139 aataaataatgactaaatagattggaagtagaagtagaagtagaagtagaagtaga 1198
Db 6780 TATATGTTTGTATTTGCTATATATATATATATATATATATATATATATATATAT
OY 1199 tgcataaataaaatcccaagttatataaatacaatcttcttcttcttcttcttct 6839
Db 6840 CAT
OY 1252 cagttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 1311
Db 6900 TATTTCTGAT
OY 1312 aacataccaactgtaaaacagaattgacaaaaaaataaaataaaataaaataaaata 1371

—

```

Yy      1650  tcttcccccctccatagatctgttctcctcttccaattctctcgtatgcgttcgcatt 1709
           ::::::::::::::::::::: : : : : : : : : : : : : : : : : :
Db      1352  yyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyyy 1411
           ::::::::::::::::::::: : : : : : : : : : : : : : : : : :

Yy      1710  cctctcgcagcccttctctccgcagcgtgttcgtcaacgcctttcggaaagc 1765
           ::::::::::::::::::::: : : : : : : : : : : : : : : : : :
Db      1412  yyyyyyyyyyyyyyyyyyygGACCAATTCCTCATCTCTTAACACTTGC 1467

RESULT      5
US-08-213-419B-3
: Sequence 3, Application US/08213419B
: Patent No. 633406
: GENERAL INFORMATION:
:   APPLICANT:   Inselburg, J. et al.
:   TITLE OF INVENTION:  GENE ENCODING PROTEIN ANTIGENS OF PLASMIDIUM FALCIPARUM
:   TITLE OF INVENTION:  AND USUS THEREFOR
:   FILE REFERENCE:  J11-002CNCBP
:   CURRENT APPLICATION NUMBER:  US/08/213, 419B
:   PRIOR FILING DATE:  1994-03-14
:   PRIOR APPLICATION NUMBER:  US 07/870,506
:   NUMBER OF SEQ ID NOS:  20
:   SOFTWARE:  Patent In Ver. 2.0
:   SEQ ID NO 3
:   LENGTH:  6124
:   TYPE:  DNA
:   ORGANISM:  Plasmodium falciparum
:   FEATURE:
:     NAME/KEY:  CDS
:     LOCATION:  (2407)..(2439)
:     NAME/KEY:  CDS
:     LOCATION:  (2598)..(3404)
:     NAME/KEY:  CDS
:     LOCATION:  (3580)..(3720)
:     NAME/KEY:  CDS
:     LOCATION:  (3850)..(5835)
:   US-08-213-419B-3

Query Match      1.6%; Score 82; DB 4; Length 6124;
Best Local Similarity 44.9%; Pred. No. 3.9e-07;
Matches 560; Conservative 0; Mismatches 670; Indels 16; Gaps 6;

Yy      159  aaaaaatgatttgggtataataattactaagaagaatttggattgtcgtgttta 218
           ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1466  aaaaaatcagctatgcttcgtaactcaaatggagataataaattatttttaa 1525
           ||||| | | | | | | | | | | | | | | | | | | | | | | | |

Yy      219  agcatatgattgtttaaactaaanaaaltgtaigttaactttaaaatgtagagta 278
           ||||| | | | | | | | | | | | | | | | | | | | | | | | |
Db      1526  ctccctagaagatatttaatgaataataatagacctatataactt-----gtatglaa 1580

Yy      279  cacatcaaatcctcgcagcatatcaaacacgcatcatagacgcagtggagaatcaaa 338
           || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1581  aacttttatcttggagaagaaaaaaatttatataatttatgtttaaattttat 1640
           ||||| | | | | | | | | | | | | | | | | | | | | | | | |

Yy      339  ga-agataatgtgattttttaaaatagatctccaaatcaatcaacttgaagaagta 397
           || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1641  tatataattatcaaatcttttttttttttttttttttttttttttttttttttttttt 1700

Yy      398  taattccttttgctcacaataaataatatatatatatatatatatataatcttgata 457
           || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1701  tatttatgaagatacaaattttttaataattatcttttttttgcgtttaaatattt---tt 1757

Yy      458  taigtcttgcaaaaaattgccaagtcnaaaaacacatgactgaactcaaacataagtcgat 517
           || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1758  tggctcagataactatccctgttatcttcttaaaaaaaatgtacacatgtacacatg 1817

Yy      518  tgaatcaaatatgaatcgatgagatgaattcccatatggttcttataacttacaac 577
           || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1818  tatatatatatatatataatgatatcttcttcaaatatttaatatatttttgaatattgaatt 1877
           || | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Page 5

```

ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/991,867B
FILING DATE: 12-DEC-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO 92/14818
FILING DATE: 12-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/827,685
FILING DATE: 30-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/657,584
FILING DATE: 19-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: DPL14.C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1511 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Amsacta moorei entemopoxvirus
FEATURE:
NAME/KEY: CDS
LOCATION: complement (18..218)
FEATURE:
NAME/KEY: CDS
LOCATION: complement (234..782)
FEATURE:
NAME/KEY: CDS
LOCATION: 852..1511
US-07-991-867B-8

```



```

Db 771 TTCATGACATTTTATATATATATATTTTCAAAAAAATTAATGCAATGA 830
Qy 784 caaataaactatccctactactactaaacatcattcatttgcacaaatcat 843
Db 831 AAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 887
Qy 844 ttcattatcatttgcatttgcatttgcatttgcatttgcatttgcatttgc 903
Db 888 ----ATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 944
Qy 904 attggttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgc 944
Db 945 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 959
Qy 960 accatttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgc 1004
Db 1005 AATATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1019
Qy 1020 ttcacagttataataataacattcacaagctatagttcttcttaaatattc 1064
Db 1065 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1079
Qy 1080 agtgcacaaagctccacataataataaagaagtagaagtagaagtagaagtag 1124
Db 1125 AAAAAATGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1139
Qy 1140 ataataatgaactaa--aaagtttgcacaaagctatagttcttcttaaatattc 1184
Db 1185 TCATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1197
Qy 1198 atgtacataaataataataataataataataataataataataataataataata 1244
Db 1245 ATATCTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1257
Qy 1258 ttcttaataatctactactaaataataataataataataataataataataata 1304
Db 1305 AATATGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1317
Qy 1318 ccaactgaataacagaatttgcacaaataataataataataataataataataata 1364
Db 1365 --AAATGATATATATCTTCAATGTTAAATAATTAATTAATTAATTAATTAATTAAT 1377
Qy 1378 aataaataccagaagacttcttgcacaaataataataataataataataataataata 1422
Db 1423 AAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1437
Qy 1438 tatcttaataataataataataataataataataataataataataataataata 1482
Db 1483 ATTTATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1504

```

```

RESULT 7
US-08-107-755A-8
Sequence 8, Application US/08107755A
Patent No. 5721352
GENERAL INFORMATION:
APPLICANT: Moyer, Richard W.
APPLICANT: Hall, Richard L.
APPLICANT: Gruidl, Michael E.
TITLE OF INVENTION: No. 5721352el Entomopoxvirus Expression System
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESS: David R. Saliwanichik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: U.S.A.
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/107,755A
FILING DATE: 19-AUG-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/827,658
FILING DATE: 30-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/657,584
FILING DATE: 19-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanichik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: 0P114.C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 372-5800
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1511 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: unknown
ORGANISM: Amsacta moorei entomopoxvirus
FEATURE:
NAME/KEY: CDS
LOCATION: complement (18..218)
NAME/KEY: CDS
LOCATION: CDS
FEATURE:
NAME/KEY: complement (234..782)
LOCATION: CDS
NAME/KEY: CDS
LOCATION: 852..1511
US-08-107-755A-8

```

```

Query Match 1.6%; Score 81.2; DB 1; Length 1511;
Best Local Similarity 45.1%; Pred. No. 4e-07;
Matches 470; Conservative 0; Mismatches 558; Indels 14; Gaps 4;
Qy 424 atatatatatatatatatatatatatatatatatatatatatatatatatatatatat 483
Db 471 ATTATTAAGCAATTTATCACAAAATGTTCTTAATTAATTAATTTCTTCAAAATGACACTC 530
Qy 484 aaaaacatgaactgaatacaataaagtcgagatgaatacaataaagtcgagatga 543
Db 531 ATCTATGCCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 590
Qy 544 ataatctcatatgcttctactacttacaacacggaataatagatataatagatata 603
Db 591 TTTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 650
Qy 604 aaagtagatctgctatataatataatataatataatataatataatataatataatata 663
Db 651 TATAAATCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 710
Qy 664 gtaattcccaataataatgaatgcggttgaataaagtcgaatgaatgaatgaatgaatga 723
Db 711 TCTATTAAATTTTCGATCAATTCCTGTTTGGCAGAAAACATTAAGACCAATTAATTAATTAAT 770
Qy 724 agttatataatcgagatcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 783
Db 771 TTCATGACATTTTATATATATATATATTTTCAAAAAAATTAATTAATTAATTAATTAATTAAT 830
Qy 784 caaataaactatccctactactactaaacatcattcatttgcacaaatcat 843
Db 831 AAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 887
Qy 844 ttcattatcatttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgcatttgc 903
Db 888 ----ATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 944

```

Mon May 13 10:17:52 2002

904 attgctttagaagaaagtagaagttgattcttcttgcaatataat---tg 959
945 atttatttnttataaaattttagttattttagaagattgcattatatttattttagattat 1004
960 acttltataataatgacttattgacacatgattacagaattacatcattacaaact 1019
1005 atttatttnttataaaattttagttattttagaagattgcattatatttattttagattat 1064
1020 ttccaaatttataataatcattcacaagacattgattcttcttcaataatttctaa 1079
1065 atttatttnttataaaattttagttattttagaagattgcattatatttattttagattat 1124
1080 agtgcacaaagactacacataataatcagaaaagtagaagttgattcttcttgca 1139
1125 aaaaattgacattttagaagatttattttagaagattgcattatatttattttagattat 1184
1140 ataaatlaattgactaa--aattgcttggaagacattgacttgaattatagaattgata 1197
1185 tcatgttaattttagaagatttattttagaagattgcattatatttattttagattat 1244
1198 atgtacataaaattttagaagatttattttagaagattgcattatatttattttagattat 1257
1245 atttatttnttataaaattttagttattttagaagattgcattatatttattttagattat 1304
1258 tctttaaataattttagaagatttattttagaagattgcattatatttattttagattat 1317
1305 aatttgcattttagaagatttattttagaagattgcattatatttattttagattat 1364
1318 ccaactgtaaaacagaaattttagaagatttattttagaagattgcattatatttattttagattat 1377
1365 --aatttgcattttagaagatttattttagaagattgcattatatttattttagattat 1422
1378 aataatcacaagagattttagaagatttattttagaagattgcattatatttattttagattat 1437
1423 aagattttagaagatttattttagaagattgcattatatttattttagattat 1482
1438 tattttaaataataaaaa 1459
1483 atttatttnttataaaattttagttattttagaagattgcattatatttattttagattat 1504

RESULT 8
US-08-544-332-8
Sequence 8, Application US/08544332
Patent No. 5935777
GENERAL INFORMATION:
APPLICANT: Moyer, Richard W.
APPLICANT: Hall, Richard L.
APPLICANT: Gruidl, Michael E.
TITLE OF INVENTION: No. 5935777 Entomopoxvirus Expression System
NUMBER OF SEQUENCES: 77
CORRESPONDENCE ADDRESS:
ADDRESS: Gerard H. Bencen
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/544.332
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/991,867
FILING DATE: 07-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/107,755

FILING DATE: 19-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO 92/14818
FILING DATE: 12-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/827,685
FILING DATE: 30-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/657,584
FILING DATE: 19-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: Bencen, Gerard H.
REGISTRATION NUMBER: 35,746
REFERENCE/DOCKET INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1511 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Amsacta moorei entomopoxvirus
FEATURE:
NAME/KEY: complement (18..218)
LOCATION: complement (234..782)
FEATURE:
NAME/KEY: complement (234..782)
LOCATION: complement (234..782)
FEATURE:
NAME/KEY: CDS
LOCATION: CDS
NAME/KEY: 852..1511
LOCATION: 852..1511
US-08-544-332-8

Query Match 1.6%; Score 81.2; DB 2; Length 1511;
Best Local Similarity 45.1%; Pred. No. 4e-07;
Matches 470; Conservative 0; Mismatches 558; Indels 14; Gaps 4;
424 at 483
471 attatttgcacatttatttatttatttatttatttatttatttatttatttatttatttattt 530
484 aaaaac 590
531 atttattgcacatttatttatttatttatttatttatttatttatttatttatttatttattt 603
544 attatttgcacatttatttatttatttatttatttatttatttatttatttatttatttattt 650
591 ttttattgcacatttatttatttatttatttatttatttatttatttatttatttatttattt 663
604 aaaaac 710
651 ttttattgcacatttatttatttatttatttatttatttatttatttatttatttatttattt 723
664 gtac 770
711 ttttattgcacatttatttatttatttatttatttatttatttatttatttatttatttattt 783
724 atttattgcacatttatttatttatttatttatttatttatttatttatttatttatttattt 830
771 ttttattgcacatttatttatttatttatttatttatttatttatttatttatttatttattt 843
784 caaaac 887
831 aaaaac 903
844 ttttattgcacatttatttatttatttatttatttatttatttatttatttatttatttattt 944
888 ---atttattgcacatttatttatttatttatttatttatttatttatttatttatttatttattt

RESULT 9
 US-07-991-867B-8/c
 Sequence No. 5476781
 Patent No. 5476781
 GENERAL INFORMATION:
 APPLICANT: Moyer, Richard W.
 APPLICANT: Hall, Richard L.
 APPLICANT: Gruidl, Michael E.
 TITLE OF INVENTION: No. 5476781a1 Entomopoxvirus
 NUMBER OF SEQUENCES: 66
 CORRESPONDENCE ADDRESS:
 ADDRESSER: David R. Saliwanich
 STREET: 2421 N.W. 41st Street, Suite A-1
 STATE: FL
 COUNTRY: USA
 ZIP: 32606
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/991.867B
 FILING DATE: 12-DEC-1992
 CLASSIFICATION: 435
 Prior APPLICATION DATA:
 APPLICATION NUMBER: WO 92/14818
 FILING DATE: 12-FEB-1992
 Prior APPLICATION DATA:
 APPLICATION NUMBER: US 07/827,685

[illegible]

```

Db 954 ANATATTAATTAATTAACATTAATCTGTTTATTAATTTTAAATTAATTAAT 895
Qy 976 ttgacttttgaacgttgatcagaattacatccacaacgttccaaagtataata 1035
Db 894 TTTAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 840
Qy 1036 aatcatcttcaagaactatagttctcttcaaatattctcaagaagatcaagactac 1095
Db 839 ATTTTATTTTCAATGATTAATTTTATTTTGAATAAATTAATTAATTAATTAAT 780
Qy 1096 cacatataatcagaataaagttagagttgattcttcttgcaataataattgactta 1155
Db 779 TCGATGAAATTAATTAATTTGCT-----CCTATGTTTTCGGCAAAACACAGAAATGATGC 725
Qy 1156 aatagtttgaagaagcacttgacttgatcattagaaatgataagtaacataaaaaatc 1215
Db 724 GAAAAATTAATTAATTAATTTTATCTAATCAAAAATGCTAATTAATTAATTAATTA 666
Qy 1216 caagttataataatcacttttccaaatgctataatcagttcttcttaaatattcact 1275
Db 667 ---TACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 612
Qy 1276 aaaaaaacactcaatataagataaattatattgaataacataccaaactgtaaaacagat 1335
Db 611 AAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 552
Qy 1336 ttgcaaaaaaataaataaataaataaataaataaataaataaataaataaataaata 1395
Db 551 TATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 494
Qy 1396 ctcttgcaaaaaatataatataacatacaataaataaataaataaataaataaataaata 1455
Db 493 TTTGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 434
Qy 1456 aaaaacagaataat 1470
Db 433 TTAACGAAATTAATTAAT 419

```

```

RESULT 10
US-08-107-755A-8/c
; Sequence 8, Application US/08107755A
; Patent No. 5721352
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruidl, Michael E.
; TITLE OF INVENTION: No. 5721352el Entomopoxvirus Expression System
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: U.S.A.
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/107,755A
; FILING DATE: 19-AUG-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/827,658
; FILING DATE: 30-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/657,584
; FILING DATE: 19-FEB-1991
; ATTORNEY/AGENT INFORMATION:

```

```

NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: 0F114.C2
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1511 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: DNA (genomic)
ORGANISM: Amsacta moorei entomopoxvirus
FEATURE:
NAME/KEY: CDS
LOCATION: complement (18..218)
FEATURE:
NAME/KEY: CDS
LOCATION: complement (234..782)
FEATURE:
NAME/KEY: CDS
LOCATION: 852..1511
US-08-107-755A-8

```

```

Query Match 1.5%; Score 79.4; DB 1; Length 1511;
Best Local Similarity 46.0%; Pred. No. 8.8e-07;
Matches 504; Conservative 0; Mismatches 566; Indels 25; Gaps 6;

```

```

Qy 376 atcaatcacttagaagaataatgtaattcttattgctcactaataaataatataat 435
Db 1488 ATTAATTTTAAATTAATTAATTAATTTCTTAATTAATTAATTAATTAATTAATTAAT 1429
Qy 436 atataataataatcttgataatagcttgacaaaatlgcgaatcaaaacactgac 495
Db 1428 TACTTTTGGAAATTTTATCAATGATTAATTTTAAATTAATTAATTAATTAATTAAT 1369
Qy 496 tgaatcaactaataagtcgaatgaatcaactaataagtcgaatgaatcaactaata 555
Db 1368 ATTTTAAATTAATTAATTAATTTCTTAATTAATTAATTAATTAATTAATTAATTAAT 1309
Qy 556 tatgtttcacttcaacttcaaaacgaaaatagataatagataatcaaaacactgatt 615
Db 1308 TATTAATCACTCACTTACTAATAGCAAAACATTAATTAATTAATTAATTAATTAAT 1253
Qy 616 ggtatataatagaagaatttggaatttcacatatacagaactaagaactccctaa 675
Db 1252 TTAATTAATTAATTAATTTTAAATTTTAAATTAATTAATTAATTAATTAATTAATTAAT 1193
Qy 676 ttaaatcagtcggttgaagaagctcaatgaatgcttgaagaattggaagaatttaaat 735
Db 1192 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1135
Qy 736 tcgagcttttttttcttgcttgctgctccaaacatttatttatttcaataataac 795
Db 1134 TGCTATTTTGTATTAAGATATATCTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1075
Qy 796 ttatccttactactaaatcattcattatcttgataatccaaacaaacatttcaatctat 855
Db 1074 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1015
Qy 856 ttgagtgctgaagaacacatcttaccagttcaaaatataataagaagtcgttcaag 915
Db 1014 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 955
Qy 916 aaaaaaagtaacagttgaattcttcttgcaaatataaattgaacttttaataataa 975
Db 954 ATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 895
Qy 976 ttgactattgaacatgattacagaatataatcaataaacttccaaagttataata 1035

```

RESULT 11
US-08-544-332-8/C
Sequence 8, Application US/08544332
Patent No. 5935777
GENERAL INFORMATION:
APPLICANT: Moyer, Richard W.
APPLICANT: Hall, Richard L.
APPLICANT: Gruidl, Michael E.
TITLE OF INVENTION: No. 5935777el Entomopoxvirus Expression System
NUMBER OF SEQUENCES: 77
CORRESPONDENCE ADDRESS:
ADDRESSEE: Gerard H. Bencen
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/544,332
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/991,867
FILING DATE: 07-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/107,755
FILING DATE: 19-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO 92/14818
FILING DATE: 12-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/827,685
FILING DATE: 30-JAN-1992

```

1 PRIOR APPLICATION DATA:
2 APPLICATION NUMBER: US 07/657,584
3 FILING DATE: 19-FEB-1991
4 ATTORNEY/AGENT INFORMATION:
5 NAME: Hengen, Gerard H.
6 REGISTRATION NUMBER: 35,746
7 REFERENCE/DOCKET NUMBER: UPL14.C4
8 TELECOMMUNICATION INFORMATION:
9 TELEPHONE: 904-375-8100
10 TELEFAX: 904-372-5800
11 INFORMATION FOR SEQ ID NO: 8:
12 SEQUENCE CHARACTERISTICS:
13 LENGTH: 1511 base pairs
14 TYPE: nucleic acid
15 STRANDEDNESS: double
16 TOPOLOGY: unknown
17 MOLECULE TYPE: DNA (genomic)
18 ORIGINAL SOURCE:
19 ORGANISM: Amsacta moorei entemopoxvirus
20 FEATURE:
21 NAME/KEY: CDS
22 LOCATION: complement (18..218)
23 FEATURE:
24 NAME/KEY: CDS
25 LOCATION: complement (234..782)
26 FEATURE:
27 NAME/KEY: CDS
28 LOCATION: 852..1511
29 US-08-544-332-8

```

[illegible]

us-09-623-514a-3.rni

Db 954 ATATATTAATTATTAATTAACATTATCTCTGTATTTATTAATTTTATTAATTAATTAATTT 895

Oy 976 ttgactatctgacaatgattccagaattaatctaccacaactctccaaagttataata 1035

Db 894 TTTAAATATTTTATTAATTTATATCGAATTTTACTTAATTCATTTGATTAATTT-----TT 840

Oy 1036 aatacattccaaagcctatctgctctctcttaaatatcttctaaagtgatccaagactac 1095

Db 839 ATTTTTTTTCATGTGATTATTTTTTTTGGAAAAAATATTCAAATTAATTAATAAAAAATG 780

Oy 1096 cacataatctcagaataaagtagaagttgattctctcttgccaataaatactgactta 1155

Db 779 TCGATAGATTAATTAATTTGGT-----CCATGTTTTCTGGCAAAACACAGATTTGATGC 725

Oy 1156 aataagtttggaagacgatgaactgattatagaaatgtgataatgataatgataaataatc 1215

Db 724 GAAAAATTAATATGATTAATTTTATTCATCAACAAAATGTGTAATTAATTAATCAATTAAT 666

Oy 1216 caagtttaataataacatcttccaaatgcatatagctatagctctctcttaaatatcttact 1275

Db 667 ----TAGATAATGATTTATTAATTAATAAATATTAATTAATCAATGACGAAATATTTTAAT 612

Oy 1276 aaaaaaacctcaaatatagataatcttcttgaaatacatccacatgtaaaacgaat 1335

Db 611 AAAAGGATTAATTTATCATTTAAACAAATTAATTTAATTAAGGAATCAATATGTAATAT 552

Oy 1336 ttgacaaaaaanaaaaaaataatgaaatggaagatgaagaacaaataaataatccacagagat 1395

Db 551 TATGATATTATTGGCTAGATGAGTGCATATTTTTTGAACAAATTCATTTAGACAA--AT 494

Oy 1396 ctatgcaaaaaaataatgatgatacacaataaacatataatgatatctttaaataaata 1455

Db 493 TTTGTGATTAATGCGCTAATTAATTAATAAAAAAGTTATTGTTCGCGAATTAATTTGACT 434

Oy 1456 aaaaacagaaaaatat 1470

Db 433 TTTACACGAATATAT 419

```

SEQUENCE CHARACTERISTICS:
LENGTH: 240 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: oligodeoxynucleotide
HYPOTHEetical: NO
ANTI-SENSE: YES
US-08-628-417-6

Query Match      1.5%; Score 77; DB 1; Length 240;
Best Local Similarity 60.1%; Pred. No. 1.7e+06; Indels 0; Gaps 0;
Matches 128; Conservative 0; Mismatches 85;

QY 1257 ctcttctaataattcacttaaaaaaacacatcatagatagaataattatgatatcacat 1316
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CTT TTTGAAATTAATTATTTACTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 80

DB 21 CTTTGAATTAATAATTTTACTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 140

QY 1317 acccaactgtaaacacagaatttgacaataaaaaaaaaaaaatgaatgaagtgaagacaa 1376
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 140

DB 81 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1436

QY 1377 aaataaatccacagagatcttatgcagaaaaaataatgatgatacacataaccatattg 1436
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

DB 141 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 200

QY 1437 atactttaataataataataaacaagaaataa 1469
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

DB 201 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 233

RESULT 13
US-09-056-075-1
Sequence 1, Application US/09056075
Patent No. 5955368
GENERAL INFORMATION:
APPLICANT: Johnson, Eric A.
APPLICANT: Bradshaw, Marite
APPLICANT: Rood, Julian
TITLE OF INVENTION: Expression System for Clostridium
TITLE OF INVENTION: Species
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESSES:
ADDRESSER: Quarles & Brady
STREET: 1 South pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,075
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296, 95238
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-9166
TELEFAX: 608-251-5000
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6243 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

```

FEATURE:
NAME/KEY: misc feature
LOCATION: 3770..4013
OTHER INFORMATION: /note= "Rp4 origin of DNA transfer (orit) from
US-09-056-075-1
Plasmid Rp4"

Query Match
Best Local Similarity 1.5%; Score 75.4; DB 2; Length 6243;
Matches 148; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 1202 acataaaatccaaagtttcaataaataacatttcaaatgctatatacgtctct 1261
DB 1139 AAAAAAGAACGATACAAAGTACCTATATACAGCGTAAAAAATGAGGCTAAAAATTA 1198
QY 1262 taataattcactaaanaaacacatcaatatagaataattatgaatacaccacaa 1321
DB 1199 AAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 1258
QY 1322 ctgtaaacagaaattgacaaaaaataaataaataaataaataaataaataaataa 1381
DB 1259 AATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAA 1318
QY 1382 aatcaccagagatctatgcaaaaaataatataatgaatacacaataacacataatg 1441
DB 1319 TAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 1378
QY 1442 tttaaaataaataaataaataaataaataaataaataaataaataaataaataa 1470
DB 1379 AATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAA 1407

RESULT 14

US-08-998-416-595/c
Sequence 595 Application US/08998416
Patent No. 6239264

GENERAL INFORMATION:
APPLICANT: Philippesen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jürgen
APPLICANT: Knechtle, Philipp
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPPII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESS: No. 6239264artis Corporation
STREET: 3054 Cornwalis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/GC01976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587

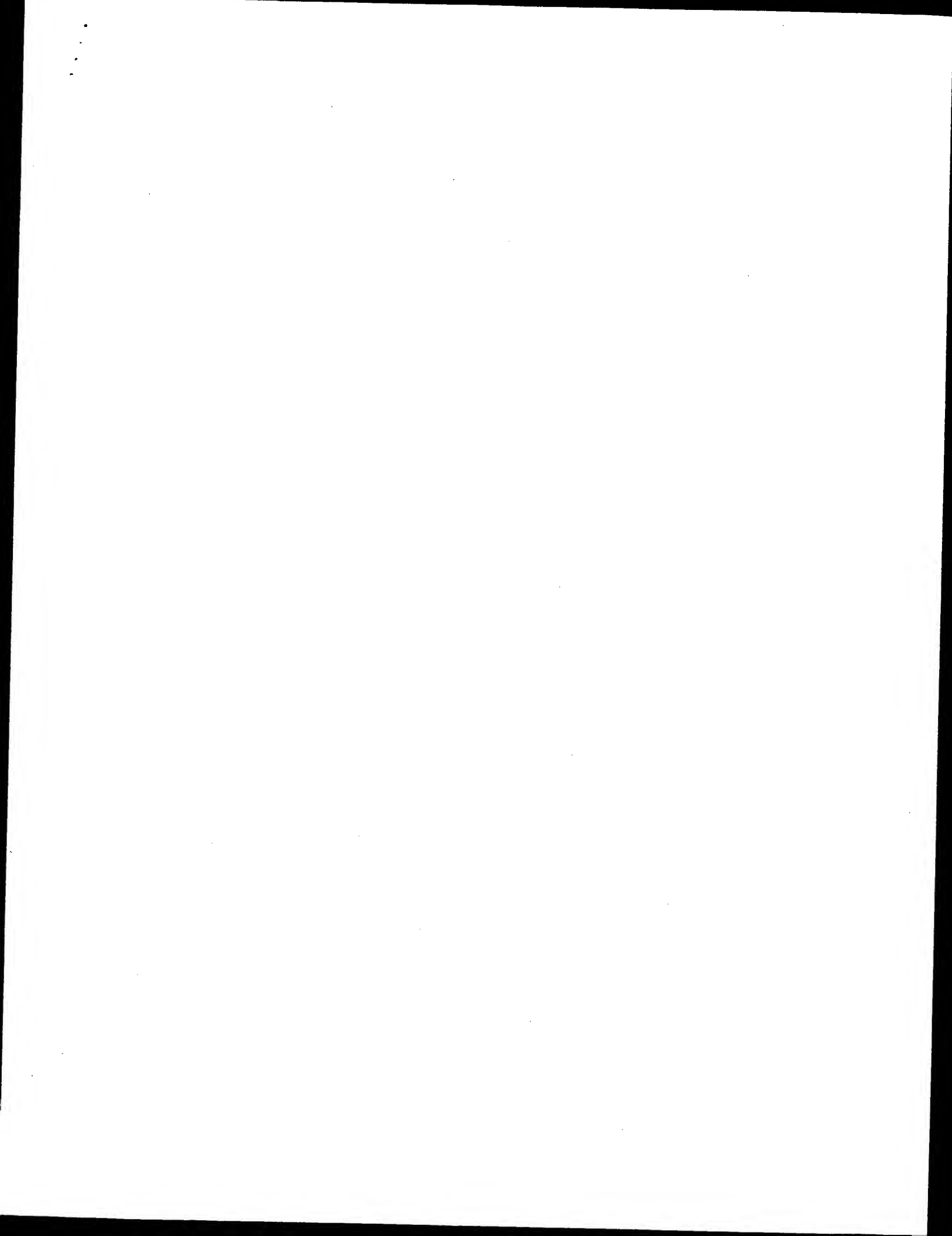
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 595:
SEQUENCE CHARACTERISTICS:
LENGTH: 658 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAC1408RP
US-08-998-416-595

Query Match
Best Local Similarity 1.4%; Score 74; DB 4; Length 658;
Matches 258; Conservative 0; Mismatches 285; Indels 3; Gaps 1;

QY 917 aaaaagttacagttgaattcttcttgcaataataaattgacttttaataat 976
DB 654 AAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 595
QY 977 tgactattgaacatgattacagaattacatcacaacattccaagttataata 1036
DB 594 AAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAA 535
QY 1037 atacattcaaaagactatagttctcttcaaaatattctcaaaagttacaaagactacc 1096
DB 534 TGAATATACAT 475
QY 1097 acataaattcagaanaagtagaagttgattcttcttgcaataataaattgacttaa 1156
DB 474 TAAATATACAT 418
QY 1157 aatgattggaagccattgactgttatagaaattgaaatgaaatgaaatgaaatgaa 1216
DB 417 AATAAATATACAT 358
QY 1217 aagttataaataacatttcaaatgctatatacagttcttcttcaataattacta 1276
DB 357 AATAAAT 298
QY 1277 aaaaacactcaataatagaataatttatttgaataacatcccaactgaaacagatt 1336
DB 297 ATAAAT 238
QY 1337 tgaacaaaaaataaataaataaataaataaataaataaataaataaataaataa 1396
DB 237 ATAAAT 178
QY 1397 ttatgcaaaaaataatataatgaataacataacataatttataataaataaataa 1456
DB 177 TATACAAACATCTAAACATATATATATATATATATATATATATATATATATATAT 112
QY 1457 aaacag 1462
DB 117 AGAGAG 112

RESULT 15

US-08-446-855A-1
Sequence 1, Application US/08446855A
Patent No. 5849573
GENERAL INFORMATION:
APPLICANT: Stewart, Thomas S
APPLICANT: Flores, Maria V
APPLICANT: O'Sullivan, William J
TITLE OF INVENTION: Nucleotide sequence encoding carbomyl
TITLE OF INVENTION: Phosphate synthetase II
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESS: Nixon & Vanderhye PC
STREET: 1100 No. 5849573th Glabe Road, 8th Floor
CITY: Arlington



FT intron /tag= g
 FT 2481..2582
 FT /tag= h
 FT 2583..2661
 FT /tag= i
 FT 2662..2758
 FT /tag= j
 FT 2759..2823
 FT /tag= k
 FT 2824..2927
 FT /tag= l
 FT 2928..3028
 FT /tag= m
 FT 3029..3108
 FT /tag= n
 FT 3109..3186
 FT /tag= o
 FT 3187..3264
 FT /tag= p
 FT 3265..3366
 FT /tag= q
 FT 3367..3466
 FT /tag= r
 FT 3467..3603
 FT /tag= s
 FT 3604..3747
 FT /tag= t
 FT 3748..3829
 FT /tag= u
 FT 3830..3933
 FT /tag= v
 FT 3934..3954
 FT /tag= w
 FT 3955..4038
 FT /tag= x
 FT 4039..4101
 FT /tag= y
 FT 4102..4184
 FT /tag= z
 FT 4185..4232
 FT /tag= aa
 FT 4233..4355
 FT /tag= ab
 FT 4356..4418
 FT /tag= ac
 FT 4419..4499
 FT /tag= ad
 FT 4500..4550
 FT /tag= ae
 FT 4551..4675
 FT /tag= af
 FT 4676..4780
 FT /tag= ag
 FT /partial
 PN MO200036114-A1.
 PD 22-JUN-2000.
 XX 16-DEC-1999; 99WO-CA01202.
 XX 17-DEC-1998; 98US-0112812.
 PA (CANVA) NAT RES COUNCIL CANADA.
 PI Zou J, Taylor DC, Wei Y, Jako CC;
 DR WPI; 2000-431592/37.
 DR P-PSDB; AAY96853.
 XX New DNA encoding diacylglycerol acyltransferase from Arabidopsis
 PT thaliana for transforming plants and regulating seed oil content, fatty
 PT acid synthesis and seed oil acyl composition in commercial and crop

PT plants
 XX Claim 2; Page 67-68; 91pp; English.
 PS
 CC This DNA encodes Arabidopsis thaliana diacylglycerol acyltransferase
 CC (DGAT). The DGAT and an insertion mutant (AS11) are useful for regulating
 CC seed oil content, the ratio of diacylglycerol/triacylglycerol proportions
 CC in seed oil, fatty acid synthesis, seed oil acyl composition, seed
 CC size/weight and carbon flux into other seed components in commercial
 CC crop plants. The natural formation of triacylglycerols can be modified to
 CC increase the yield in commercial plant oils or modify their composition
 CC to achieve specific commercial improvements of plants and plant products.
 XX Sequence 5193 BP; 1582 A; 894 C; 903 G; 1814 T; 0 other;

Query Match 100.0%; Score 5193; DB 21; Length 5193;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 5193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gctcagaccattctccgttcattgtttatttatttcaagtttaattct 60
 Db 1 gctcagaccattctccgttcattgtttatttatttcaagtttaattct 60
 QY 61 ttgtataacattcaatcttcaatgattgtgtggaaccacagatttacta 120
 Db 61 ttgtataacattcaatcttcaatgattgtgtggaaccacagatttacta 120
 QY 121 caataggaggagttactaaatagctatgctggaataatgatttattat 180
 Db 121 caataggaggagttactaaatagctatgctggaataatgatttattat 180
 QY 181 attactaaagaatatttattgttgcgtgtgttgaacatattgtttaacta 240
 Db 181 attactaaagaatatttattgttgcgtgtgttgaacatattgtttaacta 240
 QY 241 aaaaaatgattgttattcttaaaatgtaggaatcattcaaatctcgagcata 300
 Db 241 aaaaaatgattgttattcttaaaatgtaggaatcattcaaatctcgagcata 300
 QY 301 tcaaacacgattatagaccgattgagaaatagaatgaatgattttttaa 360
 Db 301 tcaaacacgattatagaccgattgagaaatagaatgaatgattttttaa 360
 QY 361 atatcgtatctcaaatcaatcattagaagaatgattcttattgttcaataa 420
 Db 361 atatcgtatctcaaatcaatcattagaagaatgattcttattgttcaataa 420
 QY 421 taatatataatataatataatataatcttattgttattgttcaataaattgcca 480
 Db 421 taatatataatataatataatataatcttattgttattgttcaataaattgcca 480
 QY 481 gtcaaaaacattgactgaatcaaatcaatgaatgattgaatcaaatgaatg 540
 Db 481 gtcaaaaacattgactgaatcaaatcaatgaatgattgaatcaaatgaatg 540
 QY 541 agtataatttccattattgttcttacttcaaacacggaataatagattatagatac 600
 Db 541 agtataatttccattattgttcttacttcaaacacggaataatagattatagatac 600
 QY 601 caaaaagtagattgtgtatattatagaagatttgaatttcaattatcagaagct 660
 Db 601 caaaaagtagattgtgtatattatagaagatttgaatttcaattatcagaagct 660
 QY 661 aagttacttccattatcaatcattgctggtgaaagtcattgaatttgaattt 720
 Db 661 aagttacttccattatcaatcattgctggtgaaagtcattgaatttgaattt 720
 QY 721 gaaatttatataatcgagatcttttttttttttttttttttttttttttt 780
 Db 721 gaaatttatataatcgagatcttttttttttttttttttttttttttttt 780
 QY 781 ttacaataatcaactatcttactaactaataatcatttattgttaccacaat 840

|||||
Db 781 ttcaaatatcaacttactactactataatcaattcatattcttgataccaacaaat 840
QY 841 cattcaatttatatttgatgtttcaagaaaacattttaccagttacaataatata 900
|||||
Db 841 cattcaatttatatttgatgtttcaagaaaacattttaccagttacaataatata 900
QY 901 aggatgtgtttgaagaaaagatacaagttgaaattctttttgtccaattataatga 960
|||||
Db 901 aggatgtgtttgaagaaaagatacaagttgaaattctttttgtccaattataatga 960
QY 961 ctttttaataataattgagcttatgaaatgattacagaaatatacttcaaaaactt 1020
Db 961 ctttttaataataattgagcttatgaaatgattacagaaatatacttcaaaaactt 1020
QY 1021 tccaagtltlaataatcatcttcaagaacttaagtcttcttcaaaaattttcaaa 1080
Db 1021 tccaagtltlaataatcatcttcaagaacttaagtcttcttcaaaaattttcaaa 1080
QY 1081 gtgataagaactacccaatataatccagaaaagtagaagttgattctttgttcaaa 1140
Db 1081 gtgataagaactacccaatataatccagaaaagtagaagttgattctttgttcaaa 1140
QY 1141 taataatgtacttaaatagtttggaaagcattgaacttgattatagaattgataatg 1200
Db 1141 taataatgtacttaaatagtttggaaagcattgaacttgattatagaattgataatg 1200
QY 1201 taacataaaaaattccaagttataataatatacttttcaaatgtctatactgtcttc 1260
Db 1201 taacataaaaaattccaagttataataatatacttttcaaatgtctatactgtcttc 1260
QY 1261 ttaaaatttctacttaaaaaaacatccaatataagaataaattatgtgaatacatcca 1320
Db 1261 ttaaaatttctacttaaaaaaacatccaatataagaataaattatgtgaatacatcca 1320
QY 1321 actgttaaaaacagaaatttgcacaaaaaaaataatgaagaatgaagaacaaaat 1380
Db 1321 actgttaaaaacagaaatttgcacaaaaaaaataatgaagaatgaagaacaaaat 1380
QY 1381 aatccacagagagatcttgcacaaaaataatagaatacaaaatccaatgtgatat 1440
Db 1381 aatccacagagagatcttgcacaaaaataatagaatacaaaatccaatgtgatat 1440
QY 1441 ttttaaaataaaaataaaaacagaaaataatcccaacacagcttttcaattaaaaacttc 1500
Db 1441 ttttaaaataaaaataaaaacagaaaataatcccaacacagcttttcaattaaaaacttc 1500
QY 1501 cgtcacacattgtgtcatcttctctctcgtcgtgaatccttttcttcttcttcttc 1560
Db 1501 cgtcacacattgtgtcatcttctctctcgtcgtgaatccttttcttcttcttcttc 1560
QY 1561 tcttcagagaaaaacttgccttcttcttcttaagaagaacagagaacagaaatccacac 1620
Db 1561 tcttcagagaaaaacttgccttcttcttcttaagaagaacagagaacagaaatccacac 1620
QY 1621 cgaatttcagcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1680
Db 1621 cgaatttcagcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1680
QY 1681 ttccaatttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 1740
Db 1681 ttccaatttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 1740
QY 1741 ttggtcaaaagcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1800
Db 1741 ttggtcaaaagcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1800
QY 1801 agaaagtgagtgagagtgatgtgcacacttgatagcttgcagagaaatcgagaatcg 1860
Db 1801 agaaagtgagtgagagtgatgtgcacacttgatagcttgcagagaaatcgagaatcg 1860
QY 1861 attcttcaaggaacttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1920
Db 1861 attcttcaaggaacttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1920
|||||
Db 1861 attcttcaaggaacttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1920
QY 1921 gggctcccgccagctttaggatacggatltgattccgttgtttaaagatgagccagagaa 1980
Db 1921 gggctcccgccagctttaggatacggatltgattccgttgtttaaagatgagccagagaa 1980
QY 1981 cagccaatttgcgcggagataataacggttgcgtgataataacagagtggttgaagagcg 2040
Db 1981 cagccaatttgcgcggagataataacggttgcgtgataataacagagtggttgaagagcg 2040
QY 2041 gcgagaaagagaaagagaaacgcgaatgcgaatgttgcgtatgcatacagacgttccagct 2100
Db 2041 gcgagaaagagaaagagaaacgcgaatgcgaatgttgcgtatgcatacagacgttccagct 2100
QY 2101 atcgagagcgagagagagatccaattagctccagacgaacttcaaaacaggtttaaactc 2160
Db 2101 atcgagagcgagagagagatccaattagctccagacgaacttcaaaacaggtttaaactc 2160
QY 2161 tcaagaacttccgaattgtgtgttgcgtgttgcgttgcgttgcgttgcgttgcgttgcgtt 2220
Db 2161 tcaagaacttccgaattgtgtgttgcgtgttgcgttgcgttgcgttgcgttgcgttgcgtt 2220
QY 2221 gtttgcattgcagagacccagccgagattatcaactcgtgtgagtagtcttattgtcgt 2280
Db 2221 gtttgcattgcagagacccagccgagattatcaactcgtgtgagtagtcttattgtcgt 2280
QY 2281 taacagtagagatcatcatccgaataatcttataagaagttgcgttgcgttgcgttgcgtt 2340
Db 2281 taacagtagagatcatcatccgaataatcttataagaagttgcgttgcgttgcgttgcgtt 2340
QY 2341 taagaattgaattgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt 2400
Db 2341 taagaattgaattgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt 2400
QY 2401 atgtgataatgtgtgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt 2460
Db 2401 atgtgataatgtgtgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt 2460
QY 2461 tggcgcttccatgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt 2520
Db 2461 tggcgcttccatgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt 2520
QY 2521 ttataagtaataatgtgtgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt 2580
Db 2521 ttataagtaataatgtgtgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt 2580
QY 2581 agtatacccttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt 2640
Db 2581 agtatacccttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt 2640
QY 2641 aatataatccagaaacgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt 2700
Db 2641 aatataatccagaaacgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt 2700
QY 2701 gaagacaagttgtatcatagaagaactcaaggttgcgttgcgttgcgttgcgttgcgttgcgtt 2760
Db 2701 gaagacaagttgtatcatagaagaactcaaggttgcgttgcgttgcgttgcgttgcgttgcgtt 2760
QY 2761 catcttcttcaataatatacaacatgcagagagttgtatccagttcaagttcaacct 2820
Db 2761 catcttcttcaataatatacaacatgcagagagttgtatccagttcaagttcaacct 2820
QY 2821 aaggtgatactgttcttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt 2880
Db 2821 aaggtgatactgttcttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt 2880
QY 2881 cccggtgatacttgaataatgcagagagttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt 2940
Db 2881 cccggtgatacttgaataatgcagagagttgcgttgcgttgcgttgcgttgcgttgcgttgcgtt 2940
QY 2941 gctcctcaactgcagatgtgtgtgtgaagttggttcttcaatgcatacagatgaatgaat 3000
Db 2941 gctcctcaactgcagatgtgtgtgtgaagttggttcttcaatgcatacagatgaatgaatgaat 3000
|||||

QY 3001 aagatccctagccaaatgcagctgataaagatacgaagaaagagctatgtattagt 3060
Db 3001 aagatccctagccaaatgcagctgataaagatacgaagaaagagctatgtattagt 3060
QY 3061 caattgcagctgattacgctgttttaacaaacactgttatgaactttagcgaacccgtga 3120
Db 3061 caattgcagctgattacgctgttttaacaaacactgttatgaactttagcgaacccgtga 3120
QY 3121 gctccctactaagcttgaagagcttgatcatattcatagctgctccacattgtgt 3180
Db 3121 gctccctactaagcttgaagagcttgatcatattcatagctgctccacattgtgt 3180
QY 3181 tataagtaactgcaaaagtgatcatcaacattcttatactgtcgaagagttctgtctaa 3240
Db 3181 tataagtaactgcaaaagtgatcatcaacattcttatactgtcgaagagttctgtctaa 3240
QY 3241 cctcgagctcttgcttcccccagcaagttatccagctgtcagatgatacgaagaggt 3300
Db 3241 cctcgagctcttgcttcccccagcaagttatccagctgtcagatgatacgaagaggt 3300
QY 3301 tgggtgctgctgaatttgtaaaactggtcatatcaacggatcatcggtattataa 3360
Db 3301 tgggtgctgctgaatttgtaaaactggtcatatcaacggatcatcggtattataa 3360
QY 3361 gaacaaagtaacttcaacatctgtcttattagtttctgttgtaaaactcatccct 3420
Db 3361 gaacaaagtaacttcaacatctgtcttattagtttctgttgtaaaactcatccct 3420
QY 3421 gctgtgtacccactgactcatgtctctgttacaatttggcagttalataacccat 3480
Db 3421 gctgtgtacccactgactcatgtctctgttacaatttggcagttalataacccat 3480
QY 3481 tgtcaggaactcaaaactccttgaagagcagatctcatatagttatgaagaggtgt 3540
Db 3481 tgtcaggaactcaaaactccttgaagagcagatctcatatagttatgaagaggtgt 3540
QY 3541 gaagcttcaggttcaaaatttatatggtgctcgtcgtacgttctactgtctctccact 3600
Db 3541 gaagcttcaggttcaaaatttatatggtgctcgtcgtacgttctactgtctctccact 3600
QY 3601 ttgtatgctgtgatacccatctcttcaaaataatttgcaaatccgaaacccgaaag 3660
Db 3601 ttgtatgctgtgatacccatctcttcaaaataatttgcaaatccgaaacccgaaag 3660
QY 3661 gctaatctcatagaatttgatatttttagttcttagagctggatgataattcaat 3720
Db 3661 gctaatctcatagaatttgatatttttagttcttagagctggatgataattcaat 3720
QY 3721 tactgaacgcaaatctctgttccaaaggttaacatatgtggcagagcttctcgtctcg 3780
Db 3721 tactgaacgcaaatctctgttccaaaggttaacatatgtggcagagcttctcgtctcg 3780
QY 3781 ggaatcgtaatttcaaaaagtgtggaatgcaaaaagtgtggagatgtgagctatt 3840
Db 3781 ggaatcgtaatttcaaaaagtgtggaatgcaaaaagtgtggagatgtgagctatt 3840
QY 3841 tactcaaaaagaaactatgaattttaatgtgtgttttttttgggtcatctaaactaac 3900
Db 3841 tactcaaaaagaaactatgaattttaatgtgtgtgttttttttgggtcatctaaactaac 3900
QY 3901 caaatcatgatttcaactgtctctcttcaactgactgagagaaigtggaattgtgtatg 3960
Db 3901 caaatcatgatttcaactgtctctcttcaactgactgagagaaigtggaattgtgtatg 3960
QY 3961 ttctcttccataaactcactcttctgtacacaaatagaagaagagcttaataaga 4020
Db 3961 ttctcttccataaactcactcttctgtacacaaatagaagaagagcttaataaga 4020
QY 4021 tctgttttcccttgacgcgtgttcaataatgaatgagtggttcaacatatctcccggtc 4080
Db 4021 tctgttttcccttgacgcgtgttcaataatgaatgagtggttcaacatatctcccggtc 4080

QY 4081 ttgcgcagcaagaataccaaagtgtgagatataacagatatagtcaattgtcgagatt 4140
Db 4081 ttgcgcagcaagaagaataccaaagtgtgagatataacagatatagtcaattgtcgagatt 4140
QY 4141 gttctgtgatalaaatttaacccctcacacactgttttttaagcactcgccatca 4200
Db 4141 gttctgtgatalaaatttaacccctcacacactgttttttaagcactcgccatca 4200
QY 4201 ttgcttccctagctctgtgaacttcttcatagagataaacttctcaatgtccgttc 4260
Db 4201 ttgcttccctagctctgtgaacttcttcatagagataaacttctcaatgtccgttc 4260
QY 4261 tctagaacgataaacaacagctagtgaaagaatgtcaattctaaacatgtttttac 4320
Db 4261 tctagaacgataaacaacagctagtgaaagaatgtcaattctaaacatgtttttac 4320
QY 4321 ttaacgactgtgtgttaaaaatttctcttcttgacagctatgcatcgagttccgttc 4380
Db 4321 ttaacgactgtgtgttaaaaatttctcttcttgacagctatgcatcgagttccgttc 4380
QY 4381 tcttaagctatgggcttctcttggaattatggttcaaggttaaaaattactaacgtc 4440
Db 4381 tcttaagctatgggcttctcttggaattatggttcaaggttaaaaattactaacgtc 4440
QY 4441 gcaatcgatttcttaactaactcaatctcatatcttgaacaaactgtttgagtag 4500
Db 4441 gcaatcgatttcttaactaactcaatctcatatcttgaacaaactgtttgagtag 4500
QY 4501 tgcctgtgttcttcaacaaactatctacaggaaggttgcgttaacaggtatgtcttc 4560
Db 4501 tgcctgtgttcttcaacaaactatctacaggaaggttgcgttaacaggtatgtcttc 4560
QY 4561 aaaaaccggaaaaatagaagaataactcttcttcaatagccatgaacttaacatcga 4620
Db 4561 aaaaaccggaaaaatagaagaataactcttcttcaatagccatgaacttaacatcga 4620
QY 4621 atgtcgaacttaataataaagtgatctgttttggaaaggatcatatatagtggtg 4680
Db 4621 atgtcgaacttaataataaagtgatctgttttggaaaggatcatatatagtggtg 4680
QY 4681 gaacatgactcttctgttcaactctcttgcgaatttggacaacccagatgtgtctctta 4740
Db 4681 gaacatgactcttctgttcaactctcttgcgaatttggacaacccagatgtgtctctta 4740
QY 4741 ttaccagacactgataaagaaagatcgaatgcataaacaactgttcaaaaatga 4800
Db 4741 ttaccagacactgataaagaaagatcgaatgcataaacaactgttcaaaaatga 4800
QY 4801 cttcttcaacatctatggtcgtctgttgatctccgttgatgtgtgtgttctatgc 4860
Db 4801 cttcttcaacatctatggtcgtctgttgatctccgttgatgtgtgtgttctatgc 4860
QY 4861 taanaacgacaatagtgatataacatgaaagaagaagaatgtgagtggtatc 4920
Db 4861 taanaacgacaatagtgatataacatgaaagaagaagaatgtgagtggtatc 4920
QY 4921 tgcataaaatttggtagagacacgcgaacccgtttgatttggatgtgtatgtaagaat 4980
Db 4921 tgcataaaatttggtagagacacgcgaacccgtttgatttggatgtgtatgtaagaat 4980
QY 4981 ttcaatcaaaaactgtgttaataatgtttacccaagaagaatgtcttctggaacagag 5040
Db 4981 ttcaatcaaaaactgtgttaataatgtttacccaagaagaatgtcttctggaacagag 5040
QY 5041 gggaaaatagtgatttctgttaggttctactgtttggaaccaatctagtaaaaactttt 5100
Db 5041 gggaaaatagtgatttctgttaggttctactgtttggaaccaatctagtaaaaactttt 5100
QY 5101 tgaataaaggaaaaaaacaaatgtaataatgcatgggagatgtgataaaccttc 5160
Db 5101 tgaataaaggaaaaaaacaaatgtaataatgcatgggagatgtgataaaccttc 5160
QY 5161 caataaagtgtatgtgtgtcccggttttgggga 5193

Db 5161 cataaagtgtgtgtcccgcttgggga 5193

Feature	Location/Qualifiers
AAAS1485	2
ID	AAA51485 standard; DNA; 5339 BP.
XX	AAA51485;
AC	
XX	09-OCT-2000 (first entry)
DF	
XX	A. thaliana As11 diacylglycerol acyltransferase genomic DNA.
DE	
XX	DGAT; diacylglycerol acyltransferase; seed oil; fatty acid synthesis;
KM	site; weight; carbon flux; TAG1; insertion mutant; ds.
XX	
OS	Arabidopsis thaliana.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	1761..4926
FT	/*tag= a
FT	1761..2150
FT	/*tag= b
FT	/partial
FT	2151..2234
FT	/*tag= c
FT	2235..2315
FT	/*tag= e
FT	2316..2380
FT	/*tag= f
FT	2370..2515
FT	/*tag= d
FT	/note= "insertion region"
FT	2381..2461
FT	/*tag= g
FT	/note= "duplication of exon 2"
FT	2462..2552
FT	/*tag= h
FT	2553..2622
FT	/*tag= i
FT	2623..2728
FT	/*tag= j
FT	2729..2804
FT	/*tag= k
FT	2805..2901
FT	/*tag= l
FT	2902..2970
FT	/*tag= m
FT	2971..3050
FT	/*tag= n
FT	3051..3174
FT	/*tag= o
FT	3175..3234
FT	/*tag= p
FT	3235..3332
FT	/*tag= q
FT	3333..3410
FT	/*tag= r
FT	3411..3512
FT	/*tag= s
FT	3513..3612
FT	/*tag= t
FT	3613..3749
FT	/*tag= u
FT	3750..3893
FT	/*tag= v
FT	3894..3975
FT	/*tag= w
FT	3976..4079
FT	/*tag= x
FT	4080..4100
FT	/*tag= y
FT	

FT	intron	4101..4184
PT		/tag= z
FT	exon	4185..4247
PT		/tag= aa
FT	intron	4248..4330
PT		/tag= aa
FT	exon	4331..4378
PT		/tag= ac
FT	intron	4379..4500
PT		/tag= ac
FT	exon	4502..4566
PT		/tag= a
FT	intron	4565..4644
PT		/tag= a
FT	exon	4646..4694
PT		/tag= a
FT	intron	4697..482
PT		/tag= a
FT	exon	4822..492
PT		/tag= a
FT		/partial

W02000036114-A1

PD 22-JUN-2000.

16-DEC-1999; 99WO-CA01202

17-DEC-1998; 98US-0112812

AA
PA (CANA) NAT RES COUNCIL CANADA.

aa
PI Zou J, Taylor DC, Wei Y, Jako CC;

AA WPI; 2000-431592/37.
DR

DR P-PSDB; AAY96854

XX New DNA encoding diacylglycerol acyltransferase from *Arabidopsis*
PT thaliana for transforming plants and regulating seed oil content, fatty
PT acid synthesis and seed oil acyl composition in commercial and crop
PT plants

XX
PS Disclosure; Page 79-81; 91pp; English

xx This DNA encodes Arabidopsis thaliana ecotype Columbia mutant AS11
cc diacylglycerol acyltransferase (DGAT). The AS11 mutant TAG1 allele has a
cc 147 bp insertion located at the central region of intron 2. The insertion
cc is a duplication of a segment that is composed of 12 bp from the 3' end
cc of intron 1, the entire sequence of exon 2 (81 bp) and 54 bp from the
cc 5' end of intron 2. The DGAT and the insertion mutant (AS11) are useful
cc for regulating seed oil content, the ratio of diacylglycerol to
cc triacylglycerol proportions in seed oil, fatty acid synthesis, seed oil
cc acyl composition, seed size/weight and carbon flux into other seed
cc components in commercial and crop plants. The natural formation of
cc triacylglycerols can be modified to increase the yield in commercial
cc plant oils or modify their composition to achieve specific commercial
cc improvements of plants and plant products.

XX Sequence 5339 BP; 1620 A; 919 C; 930 G; 1870 T; 0 other:
SQ

Query Match	97.04%	Score 5037	DB 21	Length 5539	
Best Local Similarity	97.38%	Pred. No. 0			
Matches 5193	Conservative	0	Mismatches	0	Indels 146; Gaps 1

[illegible]

QY 121 caataaggaggagtgacttaataatagctatggtgctggaataaagtattttagtataa 180
DB 121 caataaggaggagtgacttaataatagctatggtgctggaataaagtattttagtataa 180
QY 181 attatacctaagaagaattttgatttgcgtgtgttttaagcatactgattgttaactta 240
DB 181 attatacctaagaagaattttgatttgcgtgtgttttaagcatactgattgttaactta 240
QY 241 aaaaaatagtattgttaattcttaaaaatgtgagaacacatacactcgaacataa 300
DB 241 aaaaaatagtattgttaattcttaaaaatgtgagaacacatacactcgaacataa 300
QY 301 tcaaaacccgtaattatagaccgagtgtagaatcaaatagaagaataatgtatttttaaa 360
DB 301 tcaaaacccgtaattatagaccgagtgtagaatcaaatagaagaataatgtatttttaaa 360
QY 361 atatcgtatctccaatacactcacttagagaagaatgtatttctttagtgcataaaa 420
DB 361 atatcgtatctccaatacactcacttagagaagaatgtatttctttagtgcataaaa 420
QY 421 taat 480
DB 421 taat 480
QY 481 gtcaaaaacccatgactgaaacatacactaagtcggaattgaaacatacactaagtcgga 540
DB 481 gtcaaaaacccatgactgaaacatacactaagtcggaattgaaacatacactaagtcgga 540
QY 541 agtattaatttccataatgttctcacttaacttaaaaacggaaatagaatatagataac 600
DB 541 agtattaatttccataatgttctcacttaacttaaaaacggaaatagaatatagataac 600
QY 601 caaaaaagtagattgtgagataattagaagaatttgaatttcaacttcaacttcaactt 660
DB 601 caaaaaagtagattgtgagataattagaagaatttgaatttcaacttcaacttcaactt 660
QY 661 aagaatctccctaatttaattatgctgcgttgaaagaatcgaagaatgttgaatttg 720
DB 661 aagaatctccctaatttaattatgctgcgttgaaagaatcgaagaatgttgaatttg 720
QY 721 gaaagtttaatttaatttgcgactttttttgtgtgtgtgtgtgtgtgtgtgtgtgtgt 780
DB 721 gaaagtttaatttaatttgcgactttttttgtgtgtgtgtgtgtgtgtgtgtgtgtgt 780
QY 781 ttacaataatcaacttacttacttacttacttacttacttacttacttacttacttactt 840
DB 781 ttacaataatcaacttacttacttacttacttacttacttacttacttacttacttactt 840
QY 841 cattcatatcttacttacttacttacttacttacttacttacttacttacttacttactt 900
DB 841 cattcatatcttacttacttacttacttacttacttacttacttacttacttacttactt 900
QY 901 aggattgtgtttgaaataaagtaacagttgaattcttcttgaataataaattga 960
DB 901 aggattgtgtttgaaataaagtaacagttgaattcttcttgaataataaattga 960
QY 961 ctctttaaataataatgacttattgaacatgattagaataataatcacttcaactt 1020
DB 961 ctctttaaataataatgacttattgaacatgattagaataataatcacttcaactt 1020
QY 1021 tccaagtttaataataatcacttcaagactattagttcttcttaataataatcttcaaaa 1080
DB 1021 tccaagtttaataataatcacttcaagactattagttcttcttaataataatcttcaaaa 1080
QY 1081 gtgataaagactacacataatcagaagaagaatgaagaattgttcttcttgcataa 1140
DB 1081 gtgataaagactacacataatcagaagaagaatgaagaattgttcttcttgcataa 1140
QY 1141 taaataatgacttaaaatagtttggaagcgaatggaactggaatagaatgataatg 1200
DB 1141 taaataatgacttaaaatagtttggaagcgaatggaactggaatagaatgataatg 1200
QY 1201 taataaataaataatccaagtttataataataatcttcaaatgctataatcagttcttc 1260
DB 1201 taataaataaataatccaagtttataataataatcttcaaatgctataatcagttcttc 1260
QY 1261 ttaaaatattcactaaataaataaataaataaataaataaataaataaataaataaataa 1320
DB 1261 ttaaaatattcactaaataaataaataaataaataaataaataaataaataaataaataa 1320
QY 1321 actttaaagaagaatttgcacaaaaaataaataaataaataaataaataaataaataaata 1380
DB 1321 actttaaagaagaatttgcacaaaaaataaataaataaataaataaataaataaataaata 1380
QY 1381 aaatccacagaagatcttaagcaaaaaataaataaataaataaataaataaataaataa 1440
DB 1381 aaatccacagaagatcttaagcaaaaaataaataaataaataaataaataaataaataa 1440
QY 1441 ttttaaaataaataaataaataaataaataaataaataaataaataaataaataaataaata 1500
DB 1441 ttttaaaataaataaataaataaataaataaataaataaataaataaataaataaataaata 1500
QY 1501 cgtacacatgtgtatcacttctcctcgttgaaatccttcttcttcttcttcttcttctc 1560
DB 1501 cgtacacatgtgtatcacttctcctcgttgaaatccttcttcttcttcttcttcttctc 1560
QY 1561 tcttgcagaagaacttgccttcttcttcttcttcttcttcttcttcttcttcttcttctc 1620
DB 1561 tcttgcagaagaacttgccttcttcttcttcttcttcttcttcttcttcttcttcttctc 1620
QY 1621 cgaattcttagtcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctc 1680
DB 1621 cgaattcttagtcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctc 1680
QY 1681 tttcaattcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctc 1740
DB 1681 tttcaattcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctc 1740
QY 1741 ttcgtcaaacgcttcttgcgaatggcgaatttgcgaatttgcgaatttgcgaatttgcga 1800
DB 1741 ttcgtcaaacgcttcttgcgaatggcgaatttgcgaatttgcgaatttgcgaatttgcga 1800
QY 1801 aagaagtgaggaggaagtgatgcgacttgcgaatttgcgaatttgcgaatttgcgaatttgcga 1860
DB 1801 aagaagtgaggaggaagtgatgcgacttgcgaatttgcgaatttgcgaatttgcgaatttgcga 1860
QY 1861 attcttcaaggaacttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctc 1920
DB 1861 attcttcaaggaacttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctc 1920
QY 1921 gagtcccgccgagcgttaaggatcgatgcgattgcgattgcgattgcgattgcgattgcgattgcg 1980
DB 1921 gagtcccgccgagcgttaaggatcgatgcgattgcgattgcgattgcgattgcgattgcgattgcg 1980
QY 1981 cagccaatttgcgcgagataaataacggttgcgattgcgattgcgattgcgattgcgattgcg 2040
DB 1981 cagccaatttgcgcgagataaataacggttgcgattgcgattgcgattgcgattgcgattgcg 2040
QY 2041 gaggagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 2100
DB 2041 gaggaga 2100
QY 2101 atcgagggcgagagagatccacttaagtcggaagcgaatcttcaaaaggtttaaact 2160
DB 2101 atcgagggcgagagagatccacttaagtcggaagcgaatcttcaaaaggtttaaact 2160
QY 2161 tcaagaatcttcgaatttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgt 2220
DB 2161 tcaagaatcttcgaatttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgt 2220
QY 2221 gtttgcgaatgagagcgaatgcggaattatccaactcgtgtagtagttcttattgtctg 2280
DB 2221 gtttgcgaatgagagcgaatgcggaattatccaactcgtgtagtagttcttattgtctg 2280
QY 2224 ----- 2223

Db 2281 taacaagtacatcatcgcgaatactatagaagtttgctgctactgtttctcclt 2340
QY 2224 -----tgcattgcagagccctggcggattatttcaa 2254
Db 2241 taggaattgaattgcttgaanaattatcatcttgatcgagacccgcggatattcaa 2400
QY 2255 cccctgtgtagatgctctattatgcgttaaacagtagactacatcgcgaatactatgaa 2314
Db 2401 cccctgtgtagatgctctattatgcgttlaaacagtagactacatcgcgaatactatgaa 2460
QY 2315 ggttgcgttactgttctcctttaggaattgaattgcttgaanaattatcagaac 2374
Db 2461 ggttgcgttactgttctcctttaggaattgaattgcttgaanaattatcagaac 2520
QY 2375 gaataacttgttgcgttactacatcatcgaatgtagtgcgttgcgaagaacgattctc 2434
Db 2521 gaataacttgttgcgttactacatcatcgaatgtagtgcgttgcgaagaacgattctc 2580
QY 2435 ggtttagtccaagatcgctgcgagatgagccgcttccatctgctgctgtaaaagaatg 2494
Db 2581 ggtttagtccaagatcgctgcgagatgagccgcttccatctgctgctgtaaaagaatg 2640
QY 2495 ttttatttccagaagaatgatacatctgttatacgtataatgataatgtagtgcgaat 2554
Db 2641 ttttatttccagaagaatgatacatctgttatacgtataatgataatgtagtgcgaat 2700
QY 2555 cctcttgcattctctctctgttgcgagtalatcccttgcattctctcttgcgtgcct 2614
Db 2701 cctcttgcattctctctctgttgcgagtalatcccttgcattctctcttgcgtgcct 2760
QY 2615 ttaaggttgaagaattgtagtactccaagaatacatcagaacccgtgtagtaattactat 2674
Db 2761 ttaaggttgaagaattgtagtactccaagaatacatcagaacccgtgtagtaattactat 2820
QY 2675 ctcgcagcacttactgtaatttcttctgaagaagaattgttatacgaagaacttaacat 2734
Db 2821 ctcgcagcacttactgtaatttcttctgaagaagaattgttatacgaagaacttaacat 2880
QY 2735 tctgttttgaagaatgcgaaggtgtcatcttcttcttcttcttcttcttcttcttcttct 2794
Db 2881 tctgttttgaagaatgcgaaggtgtcatcttcttcttcttcttcttcttcttcttcttct 2940
QY 2795 gttttgtaccagatttaccagccctgaaggtgatactgttttctgtgtccagttgtga 2854
Db 2941 gttttgtaccagatttaccagccctgaaggtgatactgttttctgtgtccagttgtga 3000
QY 2855 tactgttttgaagtttagttgtctgcaccccggtgatacttgcgaagaatgtagaagttgattct 2914
Db 3001 tactgttttgaagtttagttgtctgcaccccggtgatacttgcgaagaatgtagaagttgattct 3060
QY 2915 gcttttttccaggtgtccacttgcagtcctcctcacttgcattgtgtgctaaagtgtgt 2974
Db 3061 gcttttttccaggtgtccacttgcagtcctcctcacttgcattgtgtgctaaagtgtgt 3120
QY 2975 tctatgtcatctagctatgatagaatcccttagcccaatgtagcgtgataagttaa 3034
Db 3121 tctatgtcatctagctatgatagaatcccttagcccaatgtagcgtgataagttaa 3180
QY 3035 tagaanaaagaacgtatgataagttacacttgcagctgttactgctttaaaccacact 3094
Db 3181 tagaanaaagaacgtatgataagttacacttgcagctgttactgctttaaaccacact 3240
QY 3095 gttatgaactttagcccaatccctgaaggtcctactactagctttagctgaagagcttggact 3154
Db 3241 gttatgaactttagcccaatccctgaaggtcctactactagctttagctgaagagcttggact 3300
QY 3155 atttcatgtgcgtcccaacattgtgttataaggttaacttgaagaatgatacacaacttct 3214
Db 3301 atttcatgtgcgtcccaacattgtgttataaggttaacttgaagaatgatacacaacttct 3360
QY 3215 atacttgaagaatgttctgtcctaaccctcgatcttgccttccccaagccaagtatc 3274
Db 3361 atacttgaagaatgttctgtcctaaccctcgatcttgccttccccaagccaagtatc 3420

QY 3275 cagcttctgcagctatcaggaaggttgggttgggtcgtcgaatttgcgaanaactgtgcat 3334
Db 3421 cagcttctgcagctatcaggaaggttgggttgggtcgtcgaatttgcgaanaactgtgcat 3480
QY 3335 tcaacgatttcaaggaatttataatagaacaagtagcgttcttcaacacttgcattactat 3394
Db 3481 tcaacgatttcaaggaatttataatagaacaagtagcgttcttcaacacttgcattactat 3540
QY 3395 tttccttgcgttaaatcatalccctgcgtgttgcacacttgcattactatgttcttct 3454
Db 3541 tttccttgcgttaaatcatalccctgcgtgttgcacacttgcattactatgttcttct 3600
QY 3455 acatttgcagctatataatccctatgtcgaagaatcgaagaacatcccttgaagaagcgt 3514
Db 3601 acatttgcagctatataatccctatgtcgaagaatcgaagaacatcccttgaagaagcgt 3660
QY 3515 ctctatacgtatcgaagaatgtagtgcgttgcagcttgcagcttgcagcttgcagcttgc 3574
Db 3661 ctctatacgtatcgaagaatgtagtgcgttgcagcttgcagcttgcagcttgcagcttgc 3720
QY 3575 tgcagttcttactgctcttccacacttgcagcttgcagcttgcagcttgcagcttgc 3634
Db 3721 tgcagttcttactgctcttccacacttgcagcttgcagcttgcagcttgcagcttgc 3780
QY 3635 ttgcgaattcgaanaaaccgaanaaagcgtataatcctcaagaatttgaattttagtt 3694
Db 3781 ttgcgaattcgaanaaaccgaanaaagcgtataatcctcaagaatttgaattttagtt 3840
QY 3695 cttaagaacggttgcagtaatttgcagttacgtacgaagaacatctcttgcgaagaat 3754
Db 3841 cttaagaacggttgcagtaatttgcagttacgtacgaagaacatctcttgcgaagaat 3900
QY 3755 atactgcagaagcttctctgcgttgcgagatcgtgaatttcaacaagaatttgcagta 3814
Db 3901 atactgcagaagcttctctgcgttgcgagatcgtgaatttcaacaagaatttgcagta 3960
QY 3815 aaaagtgtggagatgtagcatttctacccaagaanaacttgaattttagtt 3874
Db 3961 aaaagtgtggagatgtagcatttctacccaagaanaacttgaattttagtt 4020
QY 3875 cgttgcgttgcagtaatttgcagtaatttgcagtaatttgcagtaatttgcagtaatt 3934
Db 4021 cgttgcgttgcagtaatttgcagtaatttgcagtaatttgcagtaatttgcagtaatt 4080
QY 3935 actgcagaatgtaggaatgtagtgcgttcttcttcaacaacatcccttcttgcagta 3994
Db 4081 actgcagaatgtaggaatgtagtgcgttcttcttcaacaacatcccttcttgcagta 4140
QY 3995 aatagaagaaggaactaattagaatctgttcttgcagcagcgttgcagtaattgga 4054
Db 4141 aatagaagaaggaactaattagaatctgttcttgcagcagcgttgcagtaattgga 4200
QY 4055 tgggtgcagataatacttccgtgtgttgcgagcagaatgcagaaggtgtagat 4114
Db 4201 tgggtgcagataatacttccgtgtgttgcgagcagaatgcagaaggtgtagat 4260
QY 4115 ataccgatagcagatgtagatctgttcttgcagcagcagtaaccctccacacact 4174
Db 4261 ataccgatagcagatgtagatctgttcttgcagcagcagtaaccctccacacacact 4320
QY 4175 tgttctcagaacactgcagcattatcattgttcttgcagcagcagcagcagcagcagc 4234
Db 4321 tgttctcagaacactgcagcattatcattgttcttgcagcagcagcagcagcagcagc 4380
QY 4235 ataatatttcttacttgcagcagcagcagcagcagcagcagcagcagcagcagcagc 4294
Db 4381 ataatatttcttacttgcagcagcagcagcagcagcagcagcagcagcagcagcagc 4440
QY 4295 gctaatatcagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4354
Db 4441 gctaatatcagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4500

QY 4355 gctatgacgcagcttcctctgctcctcctcgaagctatgggtttcttggtatgctt 4414
 DB 4501 gctacgacgcagcttcctctgctcctcctcgaagctatgggtttcttggtatgctt 4560
 QY 4415 tcaagttaaaataactactaaactcgtcgcagctcgaattttactaaacttaactaatl 4474
 DB 4561 tcaagttaaaataactactaaactcgtcgcagctcgaattttactaaacttaactaatl 4620
 QY 4475 ctgacaacccaatttcttctggtggtggtccttctgctcctcgaagctatgggtttcttggtatgctt 4534
 DB 4621 ctgacaacccaatttcttctggtggtggtccttctgctcctcgaagctatgggtttcttggtatgctt 4680
 QY 4535 aaggttggcctcgaagctatgctcctcgaagctatgggtttcttggtatgctt 4594
 DB 4681 aaggttggcctcgaagctatgctcctcgaagctatgggtttcttggtatgctt 4740
 QY 4595 tctatgacgcagcttcctcgaagctatgggtttcttggtatgctt 4654
 DB 4741 tctatgacgcagcttcctcgaagctatgggtttcttggtatgctt 4800
 QY 4655 ggaatggatcactatcttctggtggtggtcctcgaagctatgggtttcttggtatgctt 4714
 DB 4801 ggaatggatcactatcttctggtggtggtcctcgaagctatgggtttcttggtatgctt 4860
 QY 4715 ggaacacgcagcttcctcgaagctatgggtttcttggtatgctt 4774
 DB 4861 ggaacacgcagcttcctcgaagctatgggtttcttggtatgctt 4920
 QY 4775 tcatgaaacacgcagcttcctcgaagctatgggtttcttggtatgctt 4834
 DB 4921 tcatgaaacacgcagcttcctcgaagctatgggtttcttggtatgctt 4980
 QY 4835 cgttctgctgctgctcgaagctatgggtttcttggtatgctt 4894
 DB 4981 cgttctgctgctgctcgaagctatgggtttcttggtatgctt 5040
 QY 4895 ggaacacgcagcttcctcgaagctatgggtttcttggtatgctt 4954
 DB 5041 ggaacacgcagcttcctcgaagctatgggtttcttggtatgctt 5100
 QY 4955 tggatttctgctgctcgaagctatgggtttcttggtatgctt 5014
 DB 5101 tggatttctgctgctcgaagctatgggtttcttggtatgctt 5160
 QY 5015 aaaaagaatgccttcctcgaagctatgggtttcttggtatgctt 5074
 DB 5161 aaaaagaatgccttcctcgaagctatgggtttcttggtatgctt 5220
 QY 5075 tggacacacgcagcttcctcgaagctatgggtttcttggtatgctt 5134
 DB 5221 tggacacacgcagcttcctcgaagctatgggtttcttggtatgctt 5280
 QY 5135 atgctggtggtatgctcctcgaagctatgggtttcttggtatgctt 5193
 DB 5281 atgctggtggtatgctcctcgaagctatgggtttcttggtatgctt 5339

RESULT 3
 ID AAA88835
 AC AAA88835 standard; DNA: 1942 BP.

XX 19-FEB-2001 (first entry)
 XX Arabidopsis acyl CoA:cholesterol acyltransferase DNA.
 DE Arabidopsis acyl CoA:cholesterol acyltransferase DNA.
 XX Arabidopsis acyl CoA:cholesterol acyltransferase DNA.
 KM Sterol; tocopherol; phytosterol; phytostanol; anticholesterolic;
 KM hypolipemic; ss.
 XX Arabidopsis thaliana.
 OS

PH Key Location/Qualifiers
 FT CDS 237..1799
 FT /+tag= a
 FT /FC_number= "2.3.1.26"
 PN WO200061771-A2.
 PD 19-OCT-2000.
 PF 12-APR-2000; 2000MO-US09696.
 PR 12-APR-1999; 9905-0128995.
 PA (MONS) MONSANTO CO.
 PI Venkatesh M, Corbin DR, Bhat GB, Boddupalli SS, Grebenok RJ,
 PI Kishore CM, Lardizabal KD, Lassner MW, Rangwala SH, Karunanandaa B;
 PI WPI: 2000-665136/64.
 DR P-PSDB; AAB19740.
 PS Genetically engineering the biosynthetic pathways in plants involved in
 PT the accumulation of sterol compounds and tocopherol to produce
 PT compounds for lowering the level of low density lipoprotein cholesterol
 PT in blood serum
 XX
 XX Disclosure: Page 57-58; 166pp; English.
 XX
 CC The present sequence is that of Arabidopsis thaliana DNA encoding
 CC acyl CoA:cholesterol acyltransferase (ACAT, see AAB19740). Sterol
 CC O-acetyltransferases such as ACAT catalyze the formation of
 CC cholesterol esters from cholesterol and long chain fatty acids.
 CC Recombinant constructs of the invention are used to alter the
 CC biosynthesis and accumulation of sterols and tocopherols in
 CC transgenic plants. Seeds of such plants may contain elevated
 CC levels of sitosterol and/or its esters, and alpha-tocopherol, and
 CC reduced levels of campesterol and campestanol and their esters.
 CC The seeds may also contain the novel sterol brassicasterol. Oil
 CC obtained from the seeds can be used in food and pharmaceutical
 CC compositions to lower levels of low density lipoprotein cholesterol
 CC in blood serum. ACAT enzymes can be used in the present invention
 CC to produce elevated levels of phytosterol and/or phytostanol esters.
 XX
 SQ Sequence 1942 BP; 461 A; 421 C; 425 G; 635 T; 0 other;

Query Match 12.1%; Score 626; DB 21; Length 1942;
 Best Local Similarity 100.0%; Pred. No. 7.9e-91;
 Matches 626; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1525 cttcgtgaaatccttctccttctccttctccttctcctcgaagaaacttgccttc 1584
 DB 1 cttcgtgaaatccttctccttctccttctccttctcctcgaagaaacttgccttc 60
 QY 1585 ttctataaggaaacgcagaaatccatctccacacgaattcttagcttcccttca 1644
 DB 61 ttctataaggaaacgcagaaatccatctccacacgaattcttagcttcccttca 120
 QY 1645 tccgccttccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1704
 DB 121 tccgccttccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 180
 QY 1705 cgattctcctcgaagccttcttccctcctcctcctcctcctcctcctcctcctcctc 1764
 DB 181 cgattctcctcgaagccttcttccctcctcctcctcctcctcctcctcctcctcctc 240
 QY 1765 cgatttggatctcgtcgttctactacgctggaagaaacggtggtggtggtggtggt 1824
 DB 241 cgatttggatctcgtcgttctactacgctggaagaaacggtggtggtggtggtggt 300
 QY 1825 atcttgatagcttcgtcgaagaaatcgagatcggaattcttctcgaagccttctctc 1884
 DB 301 atcttgatagcttcgtcgaagaaatcgagatcggaattcttctcgaagccttctctc 360

PI Zou J, Taylor DC, Wei Y, Jaki CC;
XX
XX WPI: 2000-431592/37.
DR P-PSDB: AAY96854.
XX
XX New DNA encoding diacylglycerol acyltransferase from Arabidopsis
PT thaliana for transforming plants and regulating seed oil content, fatty
PT acid synthesis and seed oil acyl composition in commercial and crop
PT plants

PS Claim 5: Page 79; 91pp; English.

XX This cDNA encodes Arabidopsis thaliana ecotype Columbia mutant AS11
CC diacylglycerol acyltransferase (DGAT). The AS11 mutant TAG1 allele has a
CC 147 bp insertion located at the central region of intron 2. The insertion
CC is a duplication of a segment that is composed of 12 bp from the 3' end
CC of intron 1, the entire sequence of exon 2 (81 bp) and 54 bp from the
CC 5' end of intron 2. The DGAT and the insertion mutant (AS11) are useful
CC for regulating seed oil content, the ratio of diacylglycerol to
CC triacylglycerol proportions in seed oil, fatty acid synthesis, seed oil
CC acyl composition, seed size/weight and carbon flux into other seed
CC components in commercial and crop plants. The natural formation of
CC triacylglycerols can be modified to increase the yield in commercial
CC plant oils or modify their composition to achieve specific commercial
CC improvements of plants and plant products.

XX Sequence 1985 BP; 482 A; 415 C; 446 G; 642 T; 0 other;

Query Match 11.4%; Score 592.6; DB 21; Length 1985;
Best Local Similarity 91.2%; Pred. No. 1.6e-85;
Matches 641; Conservative 0; Mismatches 59; Indels 3; Gaps 1;

OY 1623 attctttagctctctcctcaatcgcgtctctcctcctcattagattctgttctctt 1682
DB 1 attctttagctctctcctcaatcgcgtctctcctcctcattagattctgttctctt 60
OY 1683 tcaatttctctgcatgctctcgtatctctctgagcgtctcttctcccgagctgtt 1742
DB 61 tcaatttctctgcatgctctcgtatctctctcgtcgcgtctcttctcccgagctgtt 120
OY 1743 cgtcaaacgcttctcgaatgagcttctgattctcgtctgagcgtctcttctcccgag 1802
DB 121 cgtcaaacgcttctcgaatgagcttctgattctcgtctgagcgtctcttctcccgag 180
OY 1803 aacggtgagcagatctcgtatctcgtatctcgtatctcgtatctcgtatctcgtat 1862
DB 181 aacggtgagcagatctcgtatctcgtatctcgtatctcgtatctcgtatctcgtat 240
OY 1863 tcttctaacgagcttctctctctgctgctgataataatctcctcgtatgatttga 1922
DB 241 tcttctaacgagcttctctctctgctgctgataataatctcctcgtatgatttga 300
OY 1923 gttcccgagcagctttaggagatcgattgatttccgttgaacagatgagcgtctagga 1982
DB 301 gttcccgagcagctttaggagatcgattgatttccgttgaacagatgagcgtctagga 360
OY 1983 gccaatttgcgcgagataataacggtgctgagataataacggtgctgagagagcgc 2042
DB 361 gccaatttgcgcgagataataacggtgctgagataataacggtgctgagagagcgc 420
OY 2043 ggagaaagagagagagagagcagatgctgatttctgataacgagcgtcgtctcagctcat 2102
DB 421 ggagaaagagagagagagagcagatgctgatttctgataacgagcgtcgtctcagctcat 480
OY 2103 cggagagcagagagagagcagatgctgatttctgataacgagcgtcgtctcagctcat 2162
DB 481 cggagagcagagagagagcagatgctgatttctgataacgagcgtcgtctcagctcat 540
OY 2163 agaaattctgaatttggcttctgctgatttctgataacgagcgtcgtctcagctcat 2222
DB 541 ttaatacaactctgt 600

OY 2223 ttgtcattgcaagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2282
DB 601 ctt--atgaagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 657
OY 2283 aacagtagactatcatcgaataatcttgaaggtttgctgt 2325
DB 658 aacagtagactatcatcgaataatcttgaaggtttgctgt 700

RESULT 7

AAAS1482
ID AAAS1482 standard; cDNA: 1904 BP.

AC AAAS1482;

DT 09-OCT-2000 (first entry)

DE A. thaliana diacylglycerol acyltransferase cDNA.

KW DGAT: diacylglycerol acyltransferase; seed oil; fatty acid synthesis;

KW size; weight; carbon flux; ss.

XX Arabidopsis thaliana.

XX Key Location/Qualifiers

FT CDS /tag= a /product= Diacylglycerol_acyltransferase

XX W0200036114-A1.

PN 22-JUN-2000.

PD 16-DEC-1999; 99WO-CA01202.

PF 17-DEC-1998; 98US-0112812.

PR (CANADA) NAT RES COUNCIL CANADA.

PA Zou J, Taylor DC, Wei Y, Jaki CC;

PI WPI: 2000-431592/37.

DR P-PSDB: AAY96853.

XX New DNA encoding diacylglycerol acyltransferase from Arabidopsis

PT thaliana for transforming plants and regulating seed oil content, fatty

PT acid synthesis and seed oil acyl composition in commercial and crop

PT plants

PS Claim 1: Page 64-65; 91pp; English.

XX This cDNA encodes Arabidopsis thaliana diacylglycerol acyltransferase
CC (DGAT). The DGAT and an insertion mutant (AS11) are useful for regulating
CC seed oil content, the ratio of diacylglycerol/triacylglycerol proportions
CC in seed oil, fatty acid synthesis, seed oil acyl composition, seed
CC size/weight and carbon flux into other seed components in commercial and
CC crop plants. The natural formation of triacylglycerols can be modified to
CC increase the yield in commercial plant oils or modify their composition
CC to achieve specific commercial improvements of plants and plant products.

XX Sequence 1904 BP; 458 A; 399 C; 431 G; 616 T; 0 other;

Query Match 10.2%; Score 528; DB 21; Length 1904;
Best Local Similarity 100.0%; Pred. No. 3e-75;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1623 attctttagctctctcctcaatcgcgtctctcctcctcattagattctgttctctt 1682
DB 1 attctttagctctctcctcaatcgcgtctctcctcctcattagattctgttctctt 60
OY 1683 tcaatttctctgcatgctctcgtatctctctgagcgtctcttctcccgagcgtgtt 1742
DB 1 tcaatttctctgcatgctctcgtatctctctgagcgtctcttctcccgagcgtgtt 1742


```

Db      61 tcaattctctgcatgctctcagatctctcagcgcctcttctccgacgctgtt 120
QY      1743 cgtcaaacgctctcgaataggcatttgattctgctggcgcttaactaagtgagcgag 1802
Db      121 cgtcaaacgctctcgaataggcatttgattctgctggcgcttaactaagtgagcgag 1802
QY      1803 aacggtgagcgagatctgcatcttgataggctcgcagcggaatcgagatcgat 1862
Db      181 aacggtgagcgagatctgcatcttgataggctcgcagcggaatcgagatcgat 240
QY      1863 tctctaacgagactctctctctctggttcgataataatctctcctcgatgattgga 1922
Db      241 tctctaacgagactctctctctctggttcgataataatctctcctcgatgattgga 300
QY      1923 gctccgcgcagcttaaggatcgatgattcgcgtgttaacgagatcgacgagagaca 1982
Db      301 gctccgcgcagcttaaggatcgatgattcgcgtgttaacgagatcgacgagagaca 360
QY      1983 gccaatctgacgcgagataataacggtggtgagcgataacggtggtgagagcgagc 2042
Db      361 gccaatctgacgcgagataataacggtggtgagcgataacggtggtgagagcgagc 420
QY      2043 gggagagagagagagagagagcgacgagatcgatcgttgaatcgatgagagcgagc 2102
Db      421 gggagagagagagagagagagcgacgagatcgatcgttgaatcgatgagagcgagc 480
QY      2103 cggagggcgagagagagagagcgacgagatcgatcgttgaatcgatgagagcgagc 2150
Db      481 cggagggcgagagagagagagcgacgagatcgatcgttgaatcgatgagagcgagc 528

```

RESULT 8

AAA48932 ID AAA48932 standard; cDNA; 1888 BP.

AAA48932; AC

06-DEC-2000 (first entry) DT

Arabidopsis diacylglycerol acyltransferase cDNA. DE

Diacylglycerol acyltransferase; corn; rice; soybean; wheat; KM

triacylglycerol; herbicide; EC2.3.1.20; SS. XX

Arabidopsis thaliana. OS

Key Location/Qualifiers FT CDS

131..1693

/*tag= a

/product= Diacylglycerol_acyltransferase

WO200032756-A2. PN

08-JUN-2000. PD

01-DEC-1999; 99WO-US28354. PE

02-DEC-1998; 98US-0110602. PR

31-MAR-1999; 99US-0127111. PR

(DUPO) DU PONT DE NEMOURS & CO E I. PA

Cahoon EB, Kinney AJ, Cahoon RE; PI

WPI; 2000-412308/35. DR

P-PSDB; AAY94512. DR

Polynucleotides encoding diacylglycerol acetyltransferase, useful for PT

synthesis of triacylglycerols and increasing the level of oils in plant PT

seeds - XX

Claim 4; Page 38; 62pp; English. PS

XX

In the present invention, cDNA libraries from Arabidopsis, corn, rice, soybean and wheat were screened for sequences with homology to a putative acyl CoA cholesterol acyltransferase related gene from Arabidopsis thaliana and diacylglycerol acyltransferases from Homo sapiens and Mus musculus. The cDNA clones identified from this process were used to form complete diacylglycerol acyltransferase cDNA sequences. The present sequence is Arabidopsis thaliana diacylglycerol acyltransferase cDNA. Diacylglycerol acyltransferases are involved in the synthesis of triacylglycerols. Alteration of the expression of the diacylglycerol acyltransferase DNA can be useful for increasing the level of oils in plant seeds. Inhibitors of diacylglycerol acyltransferase may be useful as herbicides.

Sequence 1888 BP; 457 A; 394 C; 429 G; 608 T; 0 other;

Query Match 10.0%; Score 520; DB 21; Length 1888; Best Local Similarity 100.0%; Pred. No. 5.0e-74; Matches 520; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      1631 gcttctccttaacacgctcttccctccatagattcttctcttcaatttc 1690
Db      1 gcttctccttaacacgctcttccctccatagattcttctcttcaatttc 1690
QY      1691 tcttcagatctctgattctctctgacgctcttctcccgacgctgttcgcaaac 1750
Db      61 tcttcagatctctgattctctctgacgctcttctcccgacgctgttcgcaaac 120
QY      1751 gcttctcgaatgagatcttgattctctgagcgcttactaagtgagagagagtg 1810
Db      121 gcttctcgaatgagatcttgattctctgagcgcttactaagtgagagagagtg 180
QY      1811 cggagagtgctgcatcttgatagagcttcgcagcggaatcgagatcctcttaa 1870
Db      181 cggagagtgctgcatcttgatagagcttcgcagcggaatcgagatcctcttaa 240
QY      1871 cggactctctctctgcttgcgataataatctctctgagatgattgagatccgc 1930
Db      241 cggactctctctctgcttgcgataataatctctctgagatgattgagatccgc 300
QY      1931 cgaagtgagagatcgatgattcgttgttaacgagatcgacgagagagagcgaatt 1990
Db      301 cgaagtgagagatcgatgattcgttgttaacgagatcgacgagagagagcgaatt 360
QY      1991 ggcgcgagataacgctggtgagataacggttgagagagagcgagagagag 2050
Db      361 ggcgcgagataacgctggtgagataacggttgagagagagcgagagagag 420
QY      2051 aagagagaaagcgatgctagcttgaatcgatcgacgctgctcagctcagctcagagagc 2110
Db      421 aagagagaaagcgatgctagcttgaatcgatcgacgctgctcagctcagctcagagagc 480
QY      2111 gagagagatcactagctcgcagcaactcttcaaacag 2150
Db      481 gagagagatcactagctcgcagcaactcttcaaacag 520

```

RESULT 9

AAAC83591 ID AAAC83591 standard; DNA; 396 BP.

AAAC83591; AC

28-FEB-2001 (first entry) DT

B. napus acyltransferase partial coding sequence SEQ ID NO: 3. DE

Diacylglycerol acyltransferase; plant oil content; heart disease; KM

triacylglycerol synthesis; DAGAT; ds. XX

Brassica napus. OS

WO200066749-A1. PN

XX 09-NOV-2000.
 PD 04-MAY-2000; 2000MO-US12316.
 PF 04-MAY-1999; 99US-0132417.
 PR (CRGI) CARGILL INC.
 PA Shortosh BS;
 PI WPI: 2001-015982/02.
 DR Genes encoding diacylglycerol acyltransferase derived from Brassica
 PT napus useful for transgenic plant production with altered oil content -
 PS Claim 1; Page 15; 31pp; English.

XX The present invention provides the coding sequence for the Brassica napus
 CC diacylglycerol acyltransferase enzyme. The gene is related to that
 CC encoding the A. thaliana putative acyltransferase. The protein is
 CC involved in triacylglycerol synthesis. Oils such as triacylglycerols are
 CC a significant source of fatty acids in the human diet, contributing to
 CC illnesses such as heart disease. The gene can be used in the modulation
 CC of triacylglycerol synthesis in plants and in the production of
 CC transgenic plants which synthesise altered levels of the oil.

XX Sequence 396 BP; 95 A; 83 C; 75 G; 143 T; 0 other;
 SQ

Query Match 4.6%; Score 236.6; DB 22; Length 396;
 Best Local Similarity 81.9%; Pred. No. 5; 6e-29;
 Matches 330; Conservative 0; Mismatches 49; Indels 24; Gaps 4;

QY 2897 atgagcagtgatgattctgtttttatcaggtgctgacttgatgctcctcactgcat 2956
 Db 1 atgacagtgatgattctgtttttatcaggtgctgacttgatgctcctcactgcat 60
 QY 2957 gtgagcagtgatgattctgtttttatcaggtgctgacttgatgctcctcactgcat 3016
 Db 61 gtgagcagtgatgattctgtttttatcaggtgctgacttgatgctcctcactgcat 120
 QY 3017 gcaagcagtgatgattctgtttttatcaggtgctgacttgatgctcctcactgcat 3074
 Db 121 tcatctgataaggtataaagaagatgataa-----tatgtgctcactgcatgtgtt 173
 QY 3075 tactgttttaaccacaacactgtatgaacttaagccaatcctgaagctcctcactgcat 3134
 Db 174 tactgttttaaccacaacactgtatgaacttaagccaatcctgaagctcctcactgcat 233
 QY 3135 tagcttgaagagcttgatcattatcagtgctgctcctcactgcatgtgtt 3190
 Db 234 tagcttgaagagcttgatcattatcagtgctgctcctcactgcatgtgtt 293
 QY 3191 -----ctgcaagtgatcacaacat--tctataacttgcaagagcttctgtctaa 3239
 Db 294 atgctgctctcgtctaattgatcacaacttctcactgcaaaagtctctgtctaa 353
 QY 3240 acctcgatcttctgtcttccccaagccaagttatcaggtct 3282
 Db 354 aacctgctctcgtctaattgatcacaacttctcactgcaaaagtctctgtctaa 396

RESULT 10
 AAC83592
 ID AAC83592 standard; DNA: 410 BP.
 XX AAC83592;
 AC
 XX 28-FEB-2001 (first entry)
 DT
 XX B. napus acyltransferase partial coding sequence SEQ ID NO: 4.
 DE
 XX

KW Diacylglycerol acyltransferase; plant oil content; heart disease;
 KM triacylglycerol synthesis; DAGAT; ds.
 XX Brassica napus.
 OS
 XX WO200006749-A1.
 PN 09-NOV-2000.
 PD 04-MAY-2000; 2000MO-US12316.
 PF 04-MAY-1999; 99US-0132417.
 PR (CRGI) CARGILL INC.
 PA Shortosh BS;
 PI WPI: 2001-015982/02.
 DR Genes encoding diacylglycerol acyltransferase derived from Brassica
 PT napus useful for transgenic plant production with altered oil content -
 PS Claim 5; Page 15; 31pp; English.

XX The present invention provides the coding sequence for the Brassica napus
 CC diacylglycerol acyltransferase enzyme. The gene is related to that
 CC encoding the A. thaliana putative acyltransferase. The protein is
 CC involved in triacylglycerol synthesis. Oils such as triacylglycerols are
 CC a significant source of fatty acids in the human diet, contributing to
 CC illnesses such as heart disease. The gene can be used in the modulation
 CC of triacylglycerol synthesis in plants and in the production of
 CC transgenic plants which synthesise altered levels of the oil.

XX Sequence 410 BP; 105 A; 81 C; 62 G; 160 T; 2 other;
 SQ

Query Match 3.8%; Score 195; DB 22; Length 410;
 Best Local Similarity 78.2%; Pred. No. 2; 3e-22;
 Matches 298; Conservative 0; Mismatches 72; Indels 11; Gaps 5;

QY 3881 ttgtggtcattcactaactaaccatcagatgatactgctcctcactgcatgag 3940
 Db 29 ttgtggtcattcactaactaaccatcagatgatactgctcctcactgcatgag 88
 QY 3941 gaatgtgaataatgatactgctcctcactaaccatcctt-tttgtacacaaata 3999
 Db 89 gaatgtgaataatgatactgctcctcactaaccatcctt-tttgtacacaaata 148
 QY 4000 gaagaag--agagctaatgaatcctgttttctcctgaagcgtgttataaagatgag 4058
 Db 149 gattgaagaagaagcttattgagatctgttttctcctgaagcgtgttataaagatgag 208
 QY 4059 tcgacataatactcccggtgtgtgctgagcaagaatgagagtgagatataac 4118
 Db 209 tcgacataatactcccggtgtgtgctgagcaagaatgagagtgagatataac 268
 QY 4119 cgaatgcataatgctgagatgtgttctgtgtgataaataatgaacctcacaacgtgt 4178
 Db 269 cgaatgcataatgctgagatgtgttctgtgtgataaataatgaacctcacaacgtgt 322
 QY 4179 ttccagacatgcacatatactgttctcctcagctcctgagctcttcattagatgag 4235
 Db 323 ttccagacatgcacatatactgttctcctcagctcctgagctcttcattagatgag 382
 QY 4236 tacatacttctacatgccc 4256
 Db 383 tacatacttctacatgccc 403

RESULT 11
 ABL34155/c
 ID ABL34155 standard; DNA: 15548 BP.
 XX

QY 1407 aaatatgaatcacataaacatattgtattttaataataaacagaa 1466
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 11592 AATATATATAAATATATAAATATATAAATATATAAATATATAA 11533
 QY 1467 atatccacaccgcttccaataaat 1496
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 11532 AATATAAATATATATAAATATATAAATATATAAATATATAA 11503

RESULT 12

AAAS1486
 ID AAAS1486 standard; cDNA: 629 BP.

AC AAAS1486;

DT 09-OCT-2000 (first entry)

DE EST with homology to yeast acyl-CoA cholesterol acyltransferase.

KW DGAT; diacylglycerol acyltransferase; seed oil; fatty acid synthesis;
 size; weight; carbon flux; acyl-CoA cholesterol acyltransferase; ss.

OS Arabidopsis thaliana.

PN W0200036114-A1.

PD 22-JUN-2000.

PF 16-DEC-1999; 99WO-CA01202.

PR 17-DEC-1998; 98US-0112812.

XX (CANADA) NAT RES COUNCIL CANADA.

XX Zou J, Taylor DC, Wei Y, Jako CC;

XX WPI: 2000-431592/37.

XX New DNA encoding diacylglycerol acyltransferase from Arabidopsis

XX thaliana for transforming plants and regulating seed oil content, fatty

XX acid synthesis and seed oil acyl composition in commercial and crop

XX plants

XX Disclosure; Page 72; 91pp; English.

XX This Arabidopsis thaliana EST (expressed sequence tag) has a deduced

XX amino acid sequence showing 41 percent identity to that of the yeast

XX acyl-CoA cholesterol acyltransferase. The cDNA was used to isolate a

XX cDNA encoding A. thaliana diacylglycerol acyltransferase (DGAT). The

XX cDNA and an insertion mutant (AS11) are useful for regulating seed oil

XX content, the ratio of diacylglycerol/triacylglycerol proportions in seed

XX oil, fatty acid synthesis, seed oil acyl composition, seed size/weight

XX and carbon flux into other seed components in commercial and crop plants.

XX The natural formation of triacylglycerols can be modified to increase

XX the yield in commercial plant oils or modify their composition to achieve

XX specific commercial improvements of plants and plant products.

XX Sequence 629 BP: 150 A: 142 C: 164 G: 157 T: 16 other:

QY Query Match 2.7%; Score 141.8; DB 21; Length 629;
 Best Local Similarity 95.4%; Pred. No. 6.8e-14;
 Matches 146; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

DB 192 tctgcatgttctactgcttctccacttgg 224
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13

AAZ49453
 ID AAZ49453 standard; cDNA: 629 BP.

AC AAZ49453;

DT 04-APR-2000 (first entry)

DE A. thaliana Diacylglycerol O-Acyltransferase (DGAT) encoding cDNA.

KW Diacylglycerol O-acyltransferase; DGAT; diglyceride acyltransferase;

KW gene therapy; probe; primer; diagnosis; treatment;

KW hyperlipidemia; cardiovascular disease; obesity; diabetes; cancer;

KW neurological/immunological disorder; transgenic plant/animal; anorectic;

KW cardiant; cytostatic; antilipemic; antidiabetic; immunostimulant; ss.

OS Arabidopsis thaliana.

PN W09967403-A1.

PD 29-DEC-1999.

PF 28-AUG-1998; 98WO-US17883.

PR 24-JUN-1998; 98US-0103754.

XX (REGC) UNIV CALIFORNIA.

XX Farese RV, Cases S, Smith S, Erickson SK;

XX WPI: 2000-126632/11.

XX Novel polynucleotides and polypeptides used in treatment therapies and

XX producing of triacylglycerols

XX Claim 1: Page 47; 50pp; English.

XX The present cDNA sequence encodes A. thaliana diacylglycerol

XX O-acyltransferase (DGAT) enzyme, which is a transmembrane protein.

XX DGAT polynucleotides can be used in gene therapy, to identify DGAT

XX homologs and as source of probes and primers to identify DGAT

XX gene expression in biological specimens. DGAT protein can be used for

XX screening DGAT regulators, diagnosis and treatment of hyperlipidemia,

XX cardiovascular disease, obesity, diabetes, cancer, neurological and

XX immunological disorders. Transgenic plants with altered levels of DGAT

XX can be used to produce oils for consumption or as industrial feedstocks.

XX DGAT transgenic animals can be used as source of food and industrial

XX products, with altered triacylglyceride content.

XX Sequence 629 BP: 150 A: 142 C: 164 G: 157 T: 16 other:

QY Query Match 2.7%; Score 141.8; DB 21; Length 629;
 Best Local Similarity 95.4%; Pred. No. 6.8e-14;
 Matches 146; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3453 ttacatttggcagatataatcatttgcaggaactcaagacatccttgaagcg 3512
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 72 ttataatagaacaataataatcctattgtcaggaactcaagacatccttgaagcg 131
 QY 3513 attcttatatgctattgaagagtggtgaagcttcaagttccaatttatatgtggcg 3572
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 132 attcttatatgctattgaagagtggtgaagcttcaagttccaatttatatgtggcg 191
 QY 3573 tctgcatgttctactgcttctccacttgg 3605
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 192 tctgcatgttctactgcttctccacttgg 224


```

Db 4596 AAAAAAAAAA CTTCTTTAATAATTACTAAAAAACCTTCAATATAAAAAATAAAT 4537
OY 1304 tattgaatacactaccactgtaaaacagaatttgcaaaaaaataaaatgaatg 1363
Db 4536 AAAAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAA 4477
OY 1364 aagatgaagacaaataaatacaccagagatctatgcaaaaaatatagaatacaca 1423
Db 4476 TTAATTAATTTCAATTTAAATACCATATATATTTCAAAACAAATTAATAATATACAA 4417
OY 1424 ataaccatattgatattttaaataaaataaaacagaaa 1465
Db 4416 AAATTTAAATATATAATATAAAAAATATAAATTAATTA 4375

```

Search completed: May 13, 2002, 07:35:00
 Job time: 14584 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 07:18:25 : Search time 8260.66 Seconds
(without alignments)
13155.305 Million cell updates/sec

Title: US-09-623-514A-3
Sequence: 1 gtcacgacccattcttcc.....tggtggtccgtttgaggga 5193

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: gb_ba:*
2: gb_hgt:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hgt_hum:*
31: em_hgt_inv:*
32: em_hgt_other:*
33: em_hgtgo_inv:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	5193	100.0	92721	8	AC005917	AC005917 Arabidopsis
2	626	12.1	1942	6	AX090345	AX090345 Sequence
3	626	11.9	1942	6	AX090349	AX090349 Sequence
4	620	11.9	1988	8	ATH131831	ATH131831 Arabidops
5	551	10.6	2005	8	AY054480	AY054480 Arabidops
6	528	10.2	1904	8	ATH238008	ATH238008 Arabidops
7	477	9.2	1845	8	AF051849	AF051849 Arabidops
8	236.6	4.6	38	6	AX045795	AX045795 Sequence
9	195	3.8	410	6	AX045796	AX045796 Sequence
10	161.6	3.1	34980	6	AX344555	AX344555 Sequence
11	154	3.0	15548	6	AX347057	AX347057 Sequence
12	152.8	2.9	310721	2	AL627107	AL627107 Homo sapi
13	152.6	2.9	67970	3	PFMAL1P3	PFMAL1P3 Plasmodu
14	152.4	2.9	90487	9	AL592166	AL592166 Human DNA
15	152.4	2.9	183584	9	AC012492	AC012492 Homo sapi
16	152.2	2.9	4601	3	DMU11584	DMU11584 Drosophila
17	152.2	2.9	19517	3	DMU137541	DMU137541 Drosophila
18	147.8	2.8	181792	9	AC098822	AC098822 Homo sapi
19	147.6	2.8	172816	9	AC093899	AC093899 Homo sapi
20	146.4	2.8	159618	9	AC096750	AC096750 Homo sapi
21	146.4	2.8	199551	2	AC006281	AC006281 Plasmodu
22	146.2	2.8	86827	3	PFMAL3P5	PFMAL3P5 Plasmodu
23	144.2	2.8	326924	2	AC093082	AC093082 Homo sapi
24	144	2.8	104992	2	AC005504	AC005504 Plasmodu
25	144	2.8	169546	2	AC004157	AC004157 Plasmodu
26	143	2.8	155106	2	AC104069	AC104069 Homo sapi
27	142.8	2.7	155705	2	AC105451	AC105451 Homo sapi
28	140.8	2.7	159618	9	AC096750	AC096750 Homo sapi
29	140.4	2.7	162996	9	AC006441	AC006441 Homo sapi
30	140	2.7	86827	3	PFMAL3P5	PFMAL3P5 Plasmodu
31	139.8	2.7	172816	9	AC093899	AC093899 Homo sapi
32	138.8	2.7	143331	9	AC091214	AC091214 Homo sapi
33	138.8	2.7	184838	9	AC090740	AC090740 Homo sapi
34	138.4	2.7	144493	9	AP001547	AP001547 Homo sapi
35	138	2.7	135250	9	AC104330	AC104330 Homo sapi
36	138	2.7	181792	9	AC098822	AC098822 Homo sapi
37	137.6	2.6	14867	3	AE001398	AE001398 Plasmodu
38	137	2.6	36977	2	AC092304	AC092304 Homo sapi
39	136.8	2.6	179307	2	AC104027	AC104027 Homo sapi
40	136	2.6	36977	2	AC092304	AC092304 Homo sapi
41	136	2.6	326924	2	AC093082	AC093082 Homo sapi
42	135.8	2.6	205130	2	AC105425	AC105425 Homo sapi
43	135.4	2.6	1446	8	AF155224	AF155224 Brassica
44	135.4	2.6	1512	8	AF164434	AF164434 Brassica
45	135.2	2.6	67970	3	PFMAL1P3	PFMAL1P3 Plasmodu

ALIGNMENTS

RESULT	1	92721 bp	DNA	linear	PLN 05-APR-2000
LOCUS	AC005917				
DEFINITION	Arabidopsis thaliana chromosome II section 113 of 255 of the complete sequence. Sequence from clones F27F23, F3P11.				
ACCESSION	AC005917	AE002093			
VERSION	AC005917.2	GI:6598497			
KEYWORDS	HTG.				
SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana				
REFERENCE	1 (bases 1 to 92721)				
AUTHORS	Liu, X., Kaul, S., Rounsley, S.D., Shao, T.P., Benito, M.-I., Town, C.D., Fujii, C.Y., Mason, T.M., Bowman, C.L., Barnstead, M.E., Feldman, T.V., Buell, C.R., Ketchum, K.A., Lee, J.J., Ronning, C.M., Koo, H., Moffat, R.S., Cronin, L.A., Shen, M., VanAken, S.E., Umayam, L., Tallon, L.J., Gill, J.E., Adams, M.D., Carrera, A.J., Creasy, T.H., Goodman, H.M., Somerville, C.R., Copenhaver, G.P., Preuss, D., Nierman, W.C., White, O., Eisen, J.A., Salzberg, S.L., Fraser, C.M. and Venter, J.C.				
TITLE	Sequence and analysis of chromosome 2 of the plant Arabidopsis				

Db	301	ATCTTGATVAGGCTTCGTCACGSAATTCAGATCGGATTTCTTTAAAGCACTTCTTCTCT	360
OY	1885	ctggtttccgatatataattctctcttcggaatgagtctggaaqctcccgccgaacqytaaaggatc	1944
Db	361	CTGGTTCGCATATATATTTCTCTTCGGATGATGTTTGAGACTCCCGCGAGCTTAGGAGATC	420
OY	1945	ggatctgattccgttcttaacgaatgaacgctcaggaacacgaactttgcgcgagataata	2004
Db	421	GGATTGATTCCTGTTGTTAACGATGACGCTCAGSGAACCACCAATTTGGCCGGAGATAATA	480
OY	2005	acgggtgctgagcatataaacggttgatggaagaagcgcgagaaagaagaacacgcgcg	2064
Db	481	ACGGTGGTGCGCATATATTAACGCTGGTGGGAAGGCCGCCGAGAAAGAAAGAGAAACGCCG	540
OY	2065	abgtaacgtttaaagatcgaacgcctcggtctccagcctcatctggagggcgagaaagagatccac	2124
Db	541	ATGCTACGCTTTACGATACGACCGCTGGGTTCCAGCTCATCTGGAGGGCGAGACAGAGATCCAC	600
OY	2125	ttagctcgaagcaatcttcaaacg	2150
Db	601	TTAGCTCCGACGCAATCTTCAACAG	626

RESULT	3				
AX090349					
LOCUS					
DEFINITION	AX090349	1942 bp	DNA	linear	PAT 21-MAR-2001
ACCESSION	Sequence 42 from Patent WO0116308.				
VERSION	AX090349				
KEYWORDS	AX090349.1 GI:13444210				
SOURCE					
ORGANISM	thale cress. Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids. 1 (bases 1 to 1942) Lassner,M. and van Eenennaam,A. Plant sterol acyltransferases Patent: WO 0116308-A 42 08-MAR-2001; MONSANTO COMPANY (US)				
FEATURES					
SOURCE	location/Oualifiers 1..1942 /organism="Arabidopsis thaliana" /db_xref="taxon:3702"				
BASE COUNT	461 a 421 c 425 g 635 t				
ORIGIN					

Query Match	12.1%	Score 626;	DB 6;	Length 1942;
Best Local Similarity	100.0%	Pred. NO.	1.6e-85;	
Matches	626;	Conservative	0;	Mismatches 0;
				Indels 0;
				Gaps 0;

[illegible]

Db	301	ATCTGATAGGCTTCGTCGACGGAATTCGATCGGATTCCTTTAAACGACTTCTCT	360
QY	1885	ctgattccgataaataatttcctctggaatgagatgtggagctcccgccgaagcttaggac	1944
Db	361	CTGGTTCGATATAATTCTCTCCGAGATGATGTTGAGACTCCCGCGACGTTAGGGATC	420
QY	1945	ggattgattccgittgttlaacgagtagcgtccaggaaacccaatttgcgcgagataata	2004
Db	421	GGATTGATTCCTGTTTAACGATGACCGCTACGGAAACACCAATTTGGCCGAGATAATA	480
QY	2005	acggttgtagcagataataacggttggtgtaagagggcggcgagaaagaaagaaagccg	2064
Db	481	ACGGTGCTGCGCATATAACGGGTGGGGAAGGCGCGCGACAGAAGAAAGGAAACGCCG	540
QY	2065	atcgtaacgttaacgatacgaacgctggttccagctcatcgagagggcggaagagatccac	2124
Db	541	ATGCTAACGTTTACGATACCAACCGTGGTTCCACGCTCATGAGGAGGCGAGAGAGTCCAC	600
QY	2125	ttagctccgacgcaatcttcaacag	2150
Db	601	TTAGCTCGACCAATCTTCAACAG	626

RESULT	4
LOCUS	ATH131831
DEFINITION	1988 bp mRNA linear PLN 10-JUN-1999
ACCESSION	U01318
VERSION	U01318.1
KEYWORDS	diacylglycerol O-acyltransferase.
SOURCE	thale cress.
ORGANISM	Arabidopsis thaliana
REFERENCE	Enxarxola; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eucosids II; Brassicales; Brassicaceae; Arabidopsis.
AUTHORS	1 (bases 1 to 1988)
TITLE	Hobbs,D.H., Lu,C. and Hills,M.J.
JOURNAL	Cloning of a cDNA encoding diacylglycerol acyltransferase from
MEDLINE	Arabidopsis thaliana and its functional expression
REFERENCE	FEBS Lett. 452 (3), 145-149 (1999)
AUTHORS	2 (bases 1 to 1988)
TITLE	Hills,M.J.
JOURNAL	Direct Submission
	Submitted (22-DEC-1998) Hills M.J., Department of Brassica and

FEATURES	Location/Qualifiers
source	1. .1988

CDS

```

/organism="Arabidopsis thaliana"
/sub_species="Columbia"
/db_xref="taxon:3702"
231. 1793
/EC_number="2.3.1.20"
/codon_start=1
/product="diacylglycerol O-acyltransferase"
/protein_id="CAB44774.1"
/db_xref="GI:5050913"
/translation="MAILDSAGVTVTVENGGEFVDIDLRLRRKSRSDSSNGLLSGS
DNNSP SDVQAFADVRIDISVYNDDAQANLANLGDNNNGDNGRGGEGRGND
ATFFYRPSVPAHRAARESPLSDAIPFOSAHGFLNLCVVALINANSGLILENLRKYG
LIRDFEWSFKSLRDWPLDFMCISLISFPLAAFLVEKLVLYOKYISEPVVLFHILITM
TEVLYPVYVTLRCDSAFELSGVITMLLCIYMAIKLYSAVHLYSDYISLANADNAPV
SYVSLSKLAFEMVAPLTCOYSPRNSACIRKGNVAQPKVLIFFSGFGLIEQYIN
PIYNSHKLPLKGLDILAIERYKLSPNKLISVNLQMEYCFPHLMINILAEILCGDFREF
YKDMWNAKSVGDITNMRNMPVHKWMMVNHITLFPCLRSKIPKTLNIIAFIVSAVHELC
IADFCRLKFLMAFLGLIMEQVPLVETVTVLDERGSTGNMIFWEIFCIGQPMCVILLY
YHDLNKRKSGMS"

```


Db 301 ATTCCTCTGGATGATGTTGAGCTCCGCCGACGTTAGGATGATGATTCCTGCTG 360
 Qy 1960 ttaacgatgacgtcgaaggaacacacattgcccggagataataacgggtagcgata 2019
 Db 361 TTAACGATGACGCTCAGGAAACACGCAATTGGCCGAGATTAATACGGTGTGCGCATA 420
 Qy 2020 ataacggtggtggaagagcgcggaaggaaggaaggaacgcatgctactgact 2079
 Db 421 ATTAACGGTGTGGAAGGCGCGGAGAGAGAGAAACGCCGATGCTACGTTTACGT 480
 Qy 2080 atcgacgctggttcacgcatcagggagggagagagagtcacttaagctccgagca 2139
 Db 481 ATGACACCTCGCTCCAGCTCATCGAGGCGAGAGAGAGATCCACTTACGCTCCGACGAA 540
 Qy 2140 tcttcaaacag 2150
 Db 541 TCTTCAACAG 551

RESULT 6
 LOCUS ATH238008 1904 bp mRNA linear PLN 18-JUN-1999
 DEFINITION Arabidopsis thaliana mRNA for diacylglycerol acyltransferase.
 ACCESSION AJ238008
 VERSION AJ238008.1 GI:5123717
 KEYWORDS dgat gene; diacylglycerol acyltransferase.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 1904)
 Zou J., Wei Y., Jaki C., Selvaraj G. and Taylor D.C.
 The Arabidopsis thaliana TAC1 gene encodes for a diacylglycerol
 acyltransferase
 Unpublished
 REFERENCE 2 (bases 1 to 1904)
 Zou, J.
 JOURNAL Direct Submission
 AUTHORS Submitted (06-APR-1999) Zou J., Plant Biotechnology Institute,
 TITLE National Research Council of Canada, S7H 0W9, Saskatchewan, CANADA
 JOURNAL Location/Qualifiers
 FEATURES
 source 1..1904
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 gene 139..1701
 /gene="dgat"
 CDS 139..1701
 /gene="dgat"
 /function="lipid synthesis"
 /codon_start=1
 /product="diacylglycerol acyltransferase"
 /protein_id="CA045373.1"
 /db_xref="GI:5123718"
 /translation="MALLDSAGVTTVENGGEPVDDLRRLRRSRSDSSNGLLSSG
 DNNSPDDVGAADVRIDISVNDADGATNLANLGGNDNGGNDNGGCGEGEAGND
 ATPTVPAPARRARRESPLSDAIFKSHAGLNLGVVLIANSRLIIEENKNGM
 LIRDFPSSSLRDWPLMCCISLSEPLAFVEKLYLOKYSPEVYFLHITIM
 TEVLYPVVTRCSAFLPGVTLMLCTIWLKIVSYASYSYDRSLANADKANPEY
 SYVSKSLAYEMVAPLTCOPSPRSCATKGVARFALVLTFTGMEFTIROYIN
 PIVRSKRPLKGDLLALERLKLSEVPLIYWLDFCEFLIMLNLIAELCEGDRER
 KDKMNAKSVGDVIRMMNMPVHKMVRHLYPCPLRSKIPKTLAIIIAFLVSAVHELC
 IAVPCRLEKMAFLGIMFQYPLVFTITNLOERFSIVGNMIFFCIFGPMCVLLY
 YHDLNKRGSMS"
 BASE COUNT 458 a 399 c 431 g 616 t
 ORIGIN

Query Match 10.2%, Score 528; DB 8; Length 1904;
 Best Local Similarity 100.0%; Pred. No. 1.1e-70;
 Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1623 attcttagcttctcctcaatcgctcttctcctccatctagatctgcttctctt 1682

Db 1 ATTTCTTAGCTTCTTCTTCAATCCGCTCTTCCCTTCATTAATCTGTTCTCTCT 60
 Qy 1683 tcaattctctgacatgctctcogattctctctgaagcctcttctccagactgttt 1742
 Db 61 TCAATTTCTTCTGATGCTTCTGATTCCTCTGAGCGCTCTTTCTCCGACGCTGTTT 120
 Qy 1743 cgtcaaacgcttttgcgaatgagatcttgattctctgctgacttaacggtgagag 1802
 Db 121 CGTCAACGCTTTTGCAGAAATGCGATTTTGATTTGCTGCGGCTTACTACGCTGACGAG 180
 Qy 1803 aacggtgagagagatctgcatcttgatcttgatctgctgctgagaaatgagatcgat 1862
 Db 181 AACGTTGGCGGAGAGATGCTGATCTTGATAGCTTTCGTCAGCAAGAAATCGAGATGAT 240
 Qy 1863 tcttctaaagcacttctctctctgcttcgataataattctctctcgatgagtgtga 1922
 Db 241 TCTTCAACGACTTCTCTCTCTGTTCCGATTAATTTCTCTTCCGATGATGTTGGA 300
 Qy 1923 gctcccgacgacttagagatcgatgattatccggttgaacgatgagctcgaagaa 1982
 Db 301 GCTCCGCGCGACGTTAGGATCGGATGATTCCTGTTAACGATGACGCTCAGGAAACA 360
 Qy 1983 gccaatctgcccgaagataataacggtgtgtgagataaagcgtgtggaagagcgagc 2042
 Db 361 GCCAATTTGGCGGAGATTAATTAACGCTGTGCGCATTAATACGCTGTGGAAGAGCGCGC 420
 Qy 2043 ggaagaaggaaggaaggaagcgcgatgctgactgactgacgacgctgctcact 2102
 Db 421 GGAGAGAGGAAGAGAAACCGCATCTCAGTTTACGATCAGCGTCTCGTCCAGCTCAT 480
 Qy 2103 cgaagggcgaagagagagtcacttagctcgcagcgaactcttcaacag 2150
 Db 481 CGAGGCGGAGAGAGATCTCACTTAGCTCGGACCAATCTTCAACAG 528

RESULT 7
 LOCUS AF051849 1845 bp mRNA linear PLN 24-JAN-2000
 DEFINITION Arabidopsis thaliana diacylglycerol acyltransferase (DGAT) mRNA,
 complete cds.
 ACCESSION AF051849 AF061563
 VERSION AF051849.1 GI:6625552
 KEYWORDS
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 1845)
 Bouvier-Mave, P., Benveniste, P., Oelkers, P., Sturley, S.L. and
 Schaller, H.
 Expression in yeast and tobacco of plant cDNAs encoding acyl
 CoA:diacylglycerol acyltransferase
 Eur. J. Biochem. 267 (1), 85-96 (2000)
 20069349
 2 (bases 1 to 1845)
 Benveniste, P.
 JOURNAL Direct Submission
 MEDLINE Submitted (28-FEB-1998) Institut de Biologie Molculaire des
 PLANTES, Centre National de la Recherche Scientifique, 28 rue
 Goethe, Strasbourg 67083, France
 COMMENT On Jan 24, 2000 this sequence version replaced gi:6650202.
 FEATURES
 source 1..1845
 /organism="Arabidopsis thaliana"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /tissue_type="hypocotyl"
 /clone_lib="ABRC library, catalog number CD4-15"
 /dev_stage="3 day-old seedling"
 /note="screened with EST, GenBank Accession Number
 AA042298"

gene 1..1845
/gene="DAGAT"
CDS 88..1650
/function="catalyzes the acylation of diacylglycerol by oleoylea to give triacylglycerol"
/note="acyl-CoA cholesterol acyltransferase homolog"
/codon_start=1
/product="diacylglycerol acyltransferase"
/protein_id="AAPI9262.1"
/db_xref="GI:6625553"
/translation="MAIDSGAVTTVTENGGEFVDRLRRKSRSSNGLISGS
DNNPSDDVGAADADRIDSVNDADGCTANLADNNGGSDNGSGRGGRGNAD
ATTEYSVPARRARESPSSDAIFKSHALFNLGVVLLVANSRLIENLMKMG
LIRDEWSSRLRDPLEKMCISLISPIAAFWKLVLOKISEPVFLHIIITM
TEVLYPVYTLRCDLAFSLGVTMLTCLVIMLKIVASHTSYDISLANADKANDEV
SYVSKSLAYEMVAPTLGKOPSPASACIRGVAROFKLVIFGFGFIIEQYIN
PIVNSKRLPKGLDLAIATRVYKISYPNLYVMCMYCFEFLMNLIAELFCGPREF
YKDMNAKSVGDWRMNMVPHKMMVRHRIYFELRSLKIPKTLAIIATLYSAVFHELC
IAPCRFLKMAFLGIMFQVPLVFTITNLOERGSIVGNMIFETFCIFGPMCVLLY
YHDLNMRKGSMS"

BASE COUNT 453 a 378 c 421 g 593 t
ORIGIN

Query Match 9.2%; Score 477; DB 8; Length 1845;
Best Local Similarity 100.0%; Pred. No. 5.8e-63;
Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1674 ttctcttcaatttcttctgcatgcttctgcatctctctgagcctcttctccg 1733
DB 1 tttcccttcaatttcttctgcatgcttctgcatctctctgagcctcttctccg 60
QY 1734 aagctgttcgcaaacgcttctgcaaatgagcaattgcaattctgctgagcttactag 1793
DB 61 ACGGCTTTCGCAAAACGCTTTCGCAAAATGCGATTGCGTGGCGGCTTACTACG 120
QY 1794 gtgacgagaaagctgtgagagagctgtgcatctgataagctgtcgcagcgaagaaatcg 1853
DB 121 GTGACGGAGAACGCGTGGCGGAGAGTTCGTGATCTTGATAGGCTTCGCGACGGAATCG 180
QY 1854 agatcgatctctctcaacgagctctctctgctgcttctgcaataattctctctcgat 1913
DB 181 AGATCGGATCTCTTCAAGGAGATCTCTCTGTTGCGATTAATATTCCTTCGCGAT 240
QY 1914 gatgtgagctcccgccgagctgtaggagctgattgcttccgtgttaacagatgagct 1973
DB 241 GATGTTGAGCTCCCGCGACGCTTAGGATCGGATTGATTCGTTGTTACGATGACGCT 300
QY 1974 caggagaaacagcaatttggccgagataataacggtgggtgcaataacggtgtgga 2033
DB 301 CAGGGAACAGCCAAATTTGGCCGAGATTAATACGGTGGGAGATTAATACGGTGTGGA 360
QY 2034 aagagcgagcgagaaagaaagaaagcgcgactgctgactgctgactgactgactgact 2093
DB 361 AGAGGCGGCGAG 420
QY 2094 ccaagctcaatcgagagcgagagagagctgactgactgactgactgactgactgact 2150
DB 421 CCAGCTCATCGAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 477

RESULT 8
LOCUS AX045795 396 bp DNA linear PAT 24-NOV-2000
DEFINITION Sequence 3 from Patent WO0066749.
ACCESSION AX045795
VERSION AX045795.1 GI:11344162
KEYWORDS
SOURCE
ORGANISM
Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

REFERENCE 1 (bases 1 to 396)
AUTHORS Shortosh,B.S.
TITLE Plant acyltransferases
JOURNAL Patent: WO 0066749-A 3 09-NOV-2000;
CARGILL, INCORPORATED (US)
FEATURES location/Qualifiers
1..396
/organism="Brassica napus"
/db_xref="taxon:3708"

BASE COUNT 95 a 83 c 75 g 143 t
ORIGIN

Query Match 4.6%; Score 236.6; DB 6; Length 396;
Best Local Similarity 81.9%; Pred. No. 1.8e-26;
Matches 330; Conservative 0; Mismatches 49; Indels 24; Gaps 4;
QY 2897 atgagcaggtgcatctgcttcttcttctatcaggtgctgactgagctgctcctcactgcat 2956
DB 1 ATGACAGGTGATCTGCTCTTTATCAGGTGACGCTGATGCTGCTGCTGCTGCTGCTGCTGCT 60
QY 2957 gtgagctaaagttggttctctatgctgactgactgactgactgactgactgactgactgact 3016
DB 61 GTGTGCTGAAGTGGTCTTACGCTCATACTACTACTACTACTACTACTACTACTACTACTACT 120
QY 3017 gcaagctgaagtaagtaaacgaagaagcgtatgtaattgactgactgactgactgactgact 3074
DB 121 TCATTCGATAGGTAAGTAAGATGATATA-----TATGGTCACTGCAATGGTGT 173
QY 3075 tactgtttaacaaacactgttatagactttaagcgaactcctgagcttctctactgact 3134
DB 174 TACTATTTGACACACACACTGTGAAACGTGAGCCAACTCGAAGTCTCTACTATGT 233
QY 3135 taagctgaagagcttgacatcttccatgctgctgctccacatgctgctacagtaaa---- 3190
DB 234 TACCTGAAGAGCTTGCGATTTTCACTGCTGCTGCCACATGTGTATCAGGTAATCG 293
QY 3191 -----ctgcaagtgatcaaacat--tcttaactgcaagagttcttctctaa 3239
DB 294 ATGCGTCTTCGTAATGATATCATATCATATCTTTCACTGCAAAAGTTTCGTGCTAA 353
QY 3240 acctcgatcttgccttctcccaagcgaagttatccagcttct 3282
DB 354 AACCTTGCGTTCGTTTACCCAGCCGAGCTATCACGCTTCT 396

RESULT 9
LOCUS AX045796 410 bp DNA linear PAT 24-NOV-2000
DEFINITION Sequence 4 from Patent WO0066749.
ACCESSION AX045796
VERSION AX045796.1 GI:11344163
KEYWORDS
SOURCE
ORGANISM
Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE 1 (bases 1 to 410)
AUTHORS Shortosh,B.S.
TITLE Plant acyltransferases
JOURNAL Patent: WO 0066749-A 4 09-NOV-2000;
CARGILL, INCORPORATED (US)
FEATURES location/Qualifiers
1..410
/organism="Brassica napus"
/db_xref="taxon:3708"

BASE COUNT 105 a 81 c 62 g 160 t 2 others
ORIGIN
Query Match 3.8%; Score 195; DB 6; Length 410;

Best Local Similarity 47.1%; Pred. No. 1.5e-15;
Matches 689; Conservative 0; Mismatches 769; Indels 6; Gaps 6.

35 tattatttcaagtttaatatctccttggataacatccaacatcttcacatgaatt 94

b 214596 TATATACATAAAATATATATATATAAAATATATACATAAAATATATATAAAATATATATA 214537

95 gtgtgaaaccacagatttactacaataggggagtgacttaaatagctatgat 154

D 214536 TAAAAAATATATATAAAATATATATAATAATATAT 214477

155 gccgaaaaatgctattttagttataaattata-ctaaagaaaattttgattgtctgtt 213

214417

gccccacacgcaccccaaacctaaatgttag 273

[illegible]

333

..... 214298

11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 10

[illegible]

132

[illegible]

513 ccaattcaactataaatcccatcctccttcccccttttcctttccc

[illegible]

573 caaacqgaaataqatatataqataccaaaaaaatatat-tatatatatat+acc 631

214057 ATATATATAAATATATATATAATAATATATATAATAATATATAATAAA 213998

632 gatttggaaatttcattatcagatctaagaacttcctaattaaatcatatcatt 691

213997 TATATATAAATATATATAAATATATATATAAATATATAAATATATAAAT 213938

692 gaaaagctcaatgtaattgttgaaatttggaagttatataatcgatctttttttt 751

213937 ATATATATAAATATATATAAATATATATAAATATATATAAATATATATA 213878

52 tgglttgcgtcccaacattttattattacaataatcaacttatcctactactaa 811

[illegible]

oiz alcaatcataatccttgataccaacaatatcattcataatcttatltgatgtttaagaa 871

[illegible]

0/2 acaccacccaccagccacadaaalclalaaggaattgtgttagaaaaaaagtacaagt 931

213698

991

[illegible]

1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080 1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100 1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120 1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140 1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160 1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180 1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200 1201 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216 1217 1218 1219 1220 1221 1222 1223 1224 1225 1226 1227 1228 1229 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239 1240 1241 1242 1243 1244 1245 1246 1247 1248 1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1260 1261 1262 1263 1264 1265 1266 1267 1268 1269 1270 1271 1272 1273 1274 1275 1276 1277 1278 1279 1280 1281 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296 1297 1298 1299 1300 1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314 1315 1316 1317 1318 1319 1320 1321 1322 1323 1324 1325 1326 1327 1328 1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340 1341 1342 1343 1344 1345 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360 1361 1362 1363 1364 1365 1366 1367 1368 1369 1370 1371 1372 1373 1374 1375 1376 1377 1378 1379 1380 1381 1382 1383 1384 1385 1386 1387 1388 1389 1390 1391 1392 1393 1394 1395 1396 1397 1398 1399 1400 1401 1402 1403 1404 1405 1406 1407 1408 1409 1410 1411 1412 1413 1414 1415 1416 1417 1418 1419 1420 1421 1422 1423 1424 1425 1426 1427 1428 1429 1430 1431 1432 1433 1434 1435 1436 1437 1438 1439 1440 1441 1442 1443 1444 1445 1446 1447 1448 1449 1450 1451 1452 1453 1454 1455 1456 1457 1458 1459 1460 1461 1462 1463 1464 1465 1466 1467 1468 1469 1470 1471 1472 1473 1474 1475 1476 1477 1478 1479 1480 1481 1482 1483 1484 1485 1486 1487 1488 1489 1490 1491 1492 1493 1494 1495 1496 1497 1498 1499 1500 1501 1502 1503 1504 1505 1506 1507 1508 1509 1510 1511 1512 1513 1514 1515 1516 1517 1518 1519 1520 1521 1522 1523 1524 1525 1526 1527 1528 1529 1530 1531 1532 1533 1534 1535 1536 1537 1538 1539 1540 1541 1542 1543 1544 1545 1546 1547 1548 1549 1550 1551 1552 1553 1554 1555 1556 1557 1558 1559 1560 1561 1562 1563 1564 1565 1566 1567 1568 1569 1570 1571 1572 1573 1574 1575 1576 1577 1578 1579 1580 1581 1582 1583 1584 1585 1586 1587 1588 1589 1590 1591 1592 1593 1594 1595 1596 1597 1598 1599 1600 1601 1602 1603 1604 1605 1606 1607 1608 1609 1610 1611 1612 1613 1614 1615 1616 1617 1618 1619 1620 1621 1622 1623 1624 1625 1626 1627 1628 1629 1630 1631 1632 1633 1634 1635 1636 1637 1638 1639 1640 1641 1642 1643 1644 1645 1646 1647 1648 1649 1650 1651 1652 1653 1654 1655 1656 1657 1658 1659 1660 1661 1662 1663 1664 1665 1666 1667 1668 1669 1670 1671 1672 1673 1674 1675 1676 1677 1678 1679 1680 1681 1682 1683 1684 1685 1686 1687 1688 1689 1690 1691 1692 1693 1694 1695 1696 1697 1698 1699 1700 1701 1702 1703 1704 1705 1706 1707 1708 1709 1710 1711 1712 1713 1714 1715 1716 1717 1718 1719 1720 1721 1722 1723 1724 1725 1726 1727 1728 1729 1730 1731 1732 1733 1734 1735 1736 1737 1738 1739 1740 1741 1742 1743 1744 1745 1746 1747 1748 1749 1750 1751 1752 1753 1754 1755 1756 1757 1758 1759 1760 1761 1762 1763 1764 1765 1766 1767 1768 1769 1770 1771 1772 1773 1774 1775 1776 1777 1778 1779 1780 1781 1782 1783 1784 1785 1786 1787 1788 1789 1790 1791 1792 1793 1794 1795 1796 1797 1798 1799 1800 1801 1802 1803 1804 1805 1806 1807 1808 1809 1810 1811 1812 1813 1814 1815 1816 1817 1818 1819

100-100000

1

OY	1347	aaaaaaaaatgaaatgaagaatgaagacaaaataatcaaccagagatcttaTgcaaa	1406
Db	11649	AAATTAATAAATTAAATAAATAATATA--AAAATAAATAATTAATAAATAATTAATAAATAA	11593
OY	1407	aatatatgatcacacaataaacatcatgtgatctttttaaataaataaaaaacagaaaa	1466
Db	11592	AAATTAATAAATAATTAATAATTAATAAATAATTAATAAATAATTAATAATTAATAATTAATAA	11533
OY	1467	atacccaacacgcgccttcacatlaaat	1496
Db	11532	AAATTAATAATTAATAAATAAATAAATAAATAA	11503
RESULT	12		
LOCUS	AL627107	310721 bp	DNA linear HTG 31-JAN-2002
DEFINITION	Homo sapiens chromosome 13 clone XXYac-R12D62, *** SEQUENCING IN PROGRESS ***, 16 unordered pieces.		
ACCESSION	AL627107		
VERSION	AL627107.26	GI:18477374	
KEYWORDS	HTG; HTGS-_PHASE1; HTGS_ACTIVETRN; HTGS_DRAFT; HTGS_FULLTOP.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (stes)		
TITLE	Phillimore,B.		
JOURNAL	Direct Submission Submitted (29-JAN-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Feb 1, 2002 this sequence version replaced gi:18250812. ----- Center: Wellcome Trust Sanger Institute Center code: SC Web site: http://www.sanger.ac.uk Contact: humquery@sanger.ac.uk ----- Project Information Center project name: R12D62 ----- Summary Statistics Assembly program: XGAP4; version 4.5 Sequencing vector: plasmid; L08752; 100% of reads Chemistry: Dye-terminator Big dye; 98% of reads Chemistry: Dye-primer Big dye; 1% of reads Consensus quality: 259175 bases at least Q40 Consensus quality: 261362 bases at least Q30 Consensus quality: 262314 bases at least Q20 Insert size: 309221; sum-of-contigs Quality coverage: 5.4ix in Q20 bases; sum-of-contigs ----- NOTE: This is a 'working draft' sequence. It currently consists of 16 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved. ----- 1 2239 2238: contig of 2238 bp in length 2239 2238: gap of 100 bp 2239 4585: contig of 2247 bp in length 4586 4685: gap of 100 bp 4686 6915: contig of 2230 bp in length 6916 7015: gap of 100 bp 7016 9586: contig of 2571 bp in length 9587 9686: gap of 100 bp 9687 12002: contig of 2316 bp in length 12003 12102: gap of 100 bp 12103 15045: contig of 2943 bp in length 15046 15145: gap of 100 bp 15146 156833: contig of 14168 bp in length 156834 156933: gap of 100 bp		

FEATURES	source
*	156934 159364: contig of 2431 bp in length
*	159365 159464: gap of 100 bp
*	159465 161694: contig of 2230 bp in length
*	161695 161794: gap of 100 bp
*	161795 164473: contig of 2679 bp in length
*	164474 164573: gap of 100 bp
*	164574 166911: contig of 2338 bp in length
*	166912 167011: gap of 100 bp
*	167012 169140: contig of 2129 bp in length
*	169141 169240: gap of 100 bp
*	169241 171673: contig of 2433 bp in length
*	171674 171773: gap of 100 bp
*	171774 174200: contig of 2427 bp in length
*	174201 174300: gap of 100 bp
*	174301 176420: contig of 2120 bp in length
*	176421 176520: gap of 100 bp
*	176521 310721: contig of 134201 bp in length
	Location/Qualifiers
	I..310721

misc_feature	1. .2238
misc_feature	/note="assembly_fragment:00347"
misc_feature	2339. .4585
misc_feature	/note="assembly_fragment:00808"
misc_feature	4686. .6915
misc_feature	/note="assembly_fragment:02529"
misc_feature	7016. .9586
misc_feature	/note="assembly_fragment:02853"
misc_feature	9687. .12002
misc_feature	/note="assembly_fragment:04135"
misc_feature	12103. .15045
misc_feature	/note="assembly_fragment:04243"
misc_feature	15146. .156833
misc_feature	/note="assembly_fragment:04740"
misc_feature	156934. .159364
misc_feature	/note="assembly_fragment:04796"
misc_feature	159465. .161694
misc_feature	/note="assembly_fragment:05381"
misc_feature	161795. .164473
misc_feature	/note="assembly_fragment:05978"
misc_feature	164574. .166911
misc_feature	/note="assembly_fragment:06182"
misc_feature	167012. .169140
misc_feature	/note="assembly_fragment:06938"
misc_feature	169241. .171673
misc_feature	/note="assembly_fragment:07030"
misc_feature	171774. .174200
misc_feature	/note="assembly_fragment:07564"
misc_feature	174301. .176420
misc_feature	/note="assembly_fragment:08078"
misc_feature	176521. .310721
misc_feature	/note="assembly_fragment:08329"
BASE COUNT	86325 a 66597 c 66567 g 89326 t 1506 others
ORIGIN	

Query Match	2.98;	Score 152.8;	DB 2;	Length 310721;
Best Local Similarity	48.18;	Pred. No. 3.2e-14;		
Matches 646;	Conservative 0;	Mismatches 687;	Indels 11;	Gaps 7.
OY	134	tgaactaaatagcatgattgcgaaaaaatgtatttagtataatactacctaaga	193	
Db	218087	TAAATATATATATATATATATATATATAAATATATATATATATTAATTTAAATTAATACTTAACTAAC	218146	
OY	194	aaaatttgattgctgctgtttaagcatatgtatgttaactaaataaacatgatc	253	
Db	218147	TATTATATATTAATATATTAATAAATGCAATTTTATATATATATATATTAATATATTAATAATAT	218206	
OY	254	tgttaatcttaaataalgtgaggagtcacatc-aatactcgagcataatcaaaaccgtat	312	


```

misc_feature      complement(2984..2989)
/gene="MALP3.01"
/note="potential splice donor sequence, aaa/gtaaaa"
gene             5005..5496
/gene="MALP3.02"
5005..5496
/gene="MALP3.02"
/note="MALP3.02, hypothetical protein, len: 163 aa,
contains possible signal sequence"
/codon_start=1
/product="hypothetical protein, MALP3.02"
/protein_id="CAB63557.1"
/db_xref="GI:6594245"
/translation="MKLNREYVLCPIILFFELNSVYLGNNRRNINFHENAAK
AMKLSGETINSIKLDNGDELKIKLNDKHKDSTKMDKSYSTLSNEEYKSDTLFR
KQOEINAEANRRIIEDROEYILNNDENIATRFVLENNDELIIQSFQSLDITIQS
LN"
8020..10389
/note="possible cent, region of very high [A+T] content"
14884..20352
/gene="MALP3.03"
14884..20352
/note="MALP3.03"
/product="putative ABC transporter, len: 1822 aa"
/protein_id="CAB63558.1"
/db_xref="GI:6594246"
/translation="MTTYKENVGISKKKKRSCONISFLNLSFDWIRPLINDLIK
GDIQELPNICRNEDVPYVASKLEENLRDIEDEPYSSEKNSNEHVLKQNSDSE
KAVNYNYHNILMSILKTFKRIILISFYILETLVIGTFEIVYMLLEGOKIPV
YISFLKDFVFGSLVYVMIMPFELFEALLHFYFTLNKLSVIMLYPKIKNCISN
NHLONDAFYNTYRKPSQTEIDEISROFLSIGKNASSSGITKNNKNTDKNKEN
DYIINIKSTKMKEDSLNENRSLPNVNIIMPEDSVYTFVTSCLNLFVFKIF
MSPYVHIKISNGSCIALTSLXLSAMILFEPLSPKSYILYIRKRLDMHVL
KEKKIKMNMESPAKYINFRMKEKCKIRLILSNIGVSISSIDYEVVEFELI
YKDLRKKEELKFTSIIMPLVYKILISNANFPLVNNVMEGIVNIRKLNKYINDH
LYNDIKNTFMYRTYRNEEDYNTVDKTELOMNITSHDGTSHNKLKRLNLYINDH
MKRYEFYHKMNYHRKINIKQILSGILKVDONTNKKICFOEHRKSNSTYANNSSHE
KREYEENHNSNSTMSNEFEKKNKNEYITIKLENGSFLSGIDNKCDDHILKINEN
LRNSILAIIGVNGSKSAFESHISIDGNMTHGNYIENFRKMPILYVPONSVYFPG
RALSVEHYHMHKLCYDEKRLIQPWEIIDKDLINKNKSINYNKSGKVRCLIA
YLOKCLMDNNFYLYLDDIIFTSISKIFSNLPCKEDNISFKDSCFIISNKS
TLDNFLEIDLVQYEVANIEIDKTLKRYGINSLEYEKNLNITKESHGYSNLT
IDYTRILKLEDEVELNHKSNMKEYEAVFGKNTESYFSDISNKEYIKMKKKYK
KEHMKNNKNDNNNNNSKDDHINIMNHNRYNDINIGPSTDSPIVYSSICNE
YTIQYTSNNNSDEIYKDYKDHHEERNKSSMPFVSSSMATNPSNEKEYSSS
FRGSISETLYMPOQGVFLTSVYIPMLISFIDELKEFVLNMSIENKNEHSD
TILQKQVRYLEFVLYPLTISLVSIGCSMTIIGNTISAIYHNHNLISLINADLYF
YNNNGININRPTIDISAFDYGFKRIYKAFIFRCLISLIIIMYRDCIFLEPYV
ILLYEFVFKRESRQCEAORLYLSCHPCLNIYSNALSGKINIIYKKNYHLDAVE
HYNNFRISYEFKMLINIMASLYIKIFILLTYYIIMPHLYASGIIKRYEKRYVVI
LSTLQYCISFARLGVILKFLCDYTHKEKCCVORLEPFAKISNKNASMKNEEL
NVIQYTKERKENISDKISALVYALKNSLISINSSDDSKKRYIKRENVYSK
KITYEGRDIRTYNRKGEDSTIGILAQSSFYVYNNIRFPIDPVNPFDEIVALKIN
GINNGKNDLAIYMKQDKMSYKRIQTSKVINQSNNDNTIITNDICIRYSLVLYIN
RHKRIILIDELIPLFNQNSVHDELNSFLIGAKSFNIIIRNHPNNTVLIIISHANT
LSCDDYIVLAKGEITRYCSEYEDVKTOSELHLEMD"
23896..31533
/gene="rRNA"
/note="rRNA"
/note="region containing small subunit, 5.8S and large
subunit rRNA genes and spacer regions"
23896..31533
/gene="rRNA"
/gene="rRNA"
complement(join(31966..32476,32675..32775))
/gene="MALP3.04"
complement(join(31966..32476,32675..32775))
/gene="MALP3.04"
/note="MALP3.04, conserved hypothetical membrane protein,
len: 203 aa, similarity: P. falciparum chromosome 2,
PF01010W, O96126 predicted integral membrane protein (255

```

```

aa), fasta scores: opt: 335, E(): 4.9e-15, (36.1% identity
in 191 aa overlap)"
/product="start=1"
/codon_start=1
/product="conserved hypothetical membrane protein,
MALP3.04"
/protein_id="CAB63559.1"
/db_xref="GI:6594247"
/translation="MKSYTFINVTILLFLTLLELYYNYDFESKFKNNIKIDIN
RFRRIAEKSEEQKYPMWEDFCLINBELLRIPENSDPYLPHEYNIDKINELISNS
TKIMKEITIKMRONYEKETDMNHNMMRDPMHYKMANLYLYKVKRLINILTKDLTNI
HKEETITTKWIOEDIEYFLFNQVEMLRILLTLELFYKKE"
complement(32477..32486)
/gene="MALP3.04"
/note="potential splice acceptor sequence"
complement(32669..32674)
/gene="MALP3.04"
/note="potential splice donor sequence, aaa/gtatat"
join(36657..36743,36864..37343)
/gene="MALP3.05"
join(36657..36743,36864..37343)
/gene="MALP3.05"
/note="MALP3.05, hypothetical protein, len: 188 aa"
/codon_start=1
/product="hypothetical protein, MALP3.05"
/protein_id="CAB63560.1"
/db_xref="GI:6594248"
/translation="MRIKNSGIFRIFLLICISFTYCEFGKNCIMSYRDLIYSEN
CFNYSIDRSIAEGSSSEKRYKDIPIHELKSLINIVYEYKKEIYGVGFMMNNIN
IANVYLNIHSTFIENITFSLINSSSPVLFKFEFSIFPHMLNNVFKLIPDLCO
RNPDLHILGLIVLAIYVEENFNC"
36744..36749
/gene="MALP3.05"
/note="potential splice donor sequence, aag/gtatga"
36854..36863
/gene="MALP3.05"
/note="potential splice acceptor sequence"
complement(join(38049..39995,40210..40284))
/gene="garp"
complement(join(38049..39995,40210..40284))
/gene="MALP3.06, garp, len: 673 aa, similarity: almost
identical to GARP_PLAIP (678 aa), fasta scores: 97.6%
identity in 678 aa overlap"
/codon_start=1
/product="hypothetical garp protein"
/protein_id="CAB63561.1"
/db_xref="GI:6594249"
/translation="MNVLFSYNICILFFVCTINPTSCFNSGLKKNONILKSFDS
ITGRILNTELEKKNKDNKSKETLLKEBKEDDDVPTTSNDMLKNAHNNNELISSSTD
PTIINNNDNDNENSVKDKDKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
DKKKESEVMSLYTKGQHRPNATEHGENLYEYVSEIINNAOGLILSSPYORE
OGGCGIISVHEHSTNDTKDNKENI SEDKEDHOOEMLKTIDKREKKEKEMKOE
KIEKKRKKOEKKEKKOEKREKOEKREKOEKREKOEKREKOEKREKOEKREKOE
HKEENETEQPOOTSEETNNEIMVLPPLDYVTPREHKKKEEENKREKKEG
EKEEHEKKEEENKREKKEKREKKEKREKKEKREKKEKREKKEKREKKEKREKKE
DKQGVETIINDEKACERHITVESRPLSQPOCKILDEBQITLMDKSVKKNISTQ
EOLIGTIGVNNVPRDNKKKMAKIEBELQKQHVKEDEKKESEVEEESKEVO
EDBEVEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE
DDDAEDDDDAEDDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE
complement(39996..40005)
/gene="garp"
/note="garp"
/note="potential splice acceptor sequence"
complement(40204..40209)
/gene="garp"
/note="potential splice donor sequence, aag/gtaaca"
join(45401..46396,46562..50233)
/gene="MALP3.07"

```

```

Query Match      2.9%: Score 152.6; DB 3; Length 67970;
Best Local Similarity 47.6%: Pred. No. 4,3e-14;
Matches 618; Conservative 0; Mismatches 664; Indels 17; Gaps 5;
OY 169 tttagtataaattactaaagaataatttgctgtctgtttaagcatatgta 228

```


/note="Sequence from reads from a short insert library
derived from a clone PCR. Restriction digest data confirm
the assembly."
BASE COUNT 22057 a 22328 c 23459 g 22643 t
ORIGIN

Query Match 2.9%; Score 152.4; DB 9; Length 90487;
Best Local Similarity 46.0%; Pred. No. 4.4e-14;
Matches 662; Conservative 0; Mismatches 766; Indels 10; Gaps 4;

QY 68 aacattcaatccttcaatgttggatggtgaaaccccaagatttactacaatag 127
DB 58964 AAAATTCTATATATAATATATATAATATATAATATATAATATATAATATATA 59023
QY 128 gggagttgactaaatgcatgagtcgaaatgctttagtataatatac 187
DB 59024 AATATATATATAATATATAATATATAATATATAATATATAATATATAATATA 59083
QY 188 taagaaaatttggattgtct--gttgtaagcatatgttgaacttaaaa 244
DB 59084 TATATAAATATATATAATATATAATATATAATATATAAATATAATATATA 59143
QY 245 aatagtgttgaatccttaaaatgtgaggtacacatcaactcgcgacataca 304
DB 59144 TAAATATATATAATATATAATATATAATATATAATATATAAATATACAAATA 59203
QY 305 aaccgtatcacaagcagatgagaaatcaagaatgaatgattttaaatac 364
DB 59204 TATATAATATATACATATAATATATAATATATAATATATAAATATATAATATA 59263
QY 365 cgtatcccaaatcaatcaactcagaagataatgtaattctttagtgcataaata 424
DB 59264 TATATAATATATATAATATATAATATATAATATATAAATATATAATATATA 59323
QY 425 tatatatatatatatatatatactgttatatactgtgcaaaaatgccaagca 484
DB 59324 TAAATATATATAATATATAATATATAATATATAATATATAATATATAATATA 59382
QY 485 aaaaactgactgcaatcaactaagtcggaatgcaactataagtcgagtgta 544
DB 59383 ATATATAATATAATATAATATATAATATATAATATATAATATATAAATAATA 59442
QY 545 ttaattccatgatgttcttactacttcaacacggaatagattatagataccaa 604
DB 59443 TATAAATATATATAATATATAATATATAATATATAATATATAATATATAATATA 59502
QY 605 aagtgatgttgatataatagaagattggaatttcaatcattacagagctaaag 664
DB 59503 TAAATATAAGTATTTATGAAATATATGAAATATATAATATATAAATATATAATATA 59562
QY 665 lacttccctaataaatacatgctggttgaaaagctcaatggaatgttgaattgaaa 724
DB 59563 ATATATAATATATAATATATAATATATAAATAATATAAATAATATAATATAATATA 59622
QY 725 gttatataatcgaatccttttttgggtgtgcgccaacatttattattac 784
DB 59623 ATATATATAAATATATAAATATATAATATATAAATATATAATATATAATATATA 59682
QY 785 aata-----ataacttactactactactaataatcattcatatttgataccaca 839
DB 59683 ATATATAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAAT 59742
QY 840 tcaatcacaattctatttgatgtaagaacaacatttaccagttacaataatlat 899
DB 59743 ATATAATATATAATATATAATATATAATATATAATATATAATATATAATATA 59802
QY 900 aaggaattgttgaaaaaaagtaacagttgaattcttcttgcaaaatacaaatg 959
DB 59803 ATATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATA 59862
QY 960 accttataataatgactatgaaatgaaatgaaatgaaatgaaatgaaatgaaatg 1019

DB 59863 ATATATATAATATATAAATATATAAATATATAAATATATAAATATATAAATAT 59922
QY 1020 ttccaagttat-aataaatatattcaagaagacttagttcttcaaatatttciaa 1078
DB 59923 ATACAAATATATAATATAATATATAATATAAATATATAAATATATAAATATATA 59982
QY 1079 aagtgatcagaagactcaccataataatcagaagaagtagaagttgattcttctgta 1138
DB 59983 ATATATAATATATAAATATATAAATATATAAATATATAAATATATAAATATATA 60042
QY 1139 aataataatgacttaaatagttggaaagccatgaaacttgattgattagattgataa 1198
DB 60043 TATAAATGATATATAATATATAAATATATAAATATATAAATATATAAATATATAAC 60102
QY 1199 tgcataaaaaaattccaagtttaataatcatttttcaaatgctatacagttct 1258
DB 60103 AATATATATAATATATAAATATATAAATATATAAATATATAAATATATAAATATA 60162
QY 1259 tcttaaatatttcaataaaaaaacctcaaatatagaataatttattgaatacatcac 1318
DB 60163 TAACAATATATATAATATATAATATATAATATATAAATATATAAATATATAAATATA 60222
QY 1319 caactgtaaaacaggaatttgcaaaaaaataaaatgaatgaagatgaagacaata 1378
DB 60223 ATAAATATATATAATATATAATATATAATATATAAATATATAAATATATAATATA 60282
QY 1379 ataactcagcagagatcttgcataaaatataatgataacacataatgattat 1438
DB 60283 ATATATATAATATATAAATATATAAATATATAAATATATAAATATATAAATATA 60342
QY 1439 attttaaataataaaaaacagaataataatcccaacccgcttttcaataaaat 1496
DB 60343 ATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATA 60400

RESULT 15
AC012492/c 183584 bp DNA linear PRI 07-NOV-2001
LOCUS
DEFINITION
AC012492
ACCESSION
VERSION
KEYWORDS
HTG.
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 183584)
Sulston,J.E. and Waterston,R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
2 (bases 1 to 183584)
Belter,E., Haekenson,W. and Lewis,S.
The sequence of Homo sapiens BAC clone RP11-224B10
unpublished
3 (bases 1 to 183584)
Waterston,R.H.
Direct Submission
Submitted (28-OCT-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 183584)
Waterston,R.H.
Direct Submission
Submitted (20-MAY-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 183584)
Waterston,R.
Direct Submission
Submitted (07-NOV-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 20, 2001 this sequence version replaced gi:13431121.

----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: saplens@wustl.wustl.edu
 Summary Statistics
 Center project name: H_NH0224B10

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION: Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION: The Rpcl1-11 human BAC library was made from the blood of one male donor, as described by Oseagawa, K., Moon, P.Y., Zhao, B., Frenken, E., Tatenno, M., Caranese, J.J. and de Jong, P.J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION: The clone sequenced to the left is Rp11-62P16. Actual start of this clone is at base position 1 of Rp11-224B10; actual end is at base position 183584 of Rp11-224B10.

The fidelity of the sequence from base position 14880 to 16471 cannot be guaranteed due to low quality data within a tandem repeat.

There are polymorphic base differences in the overlap of RP11-224B10 with RP11-62P16 as well as with the redundant clone CTD-2021011.

Data from CTD-2021011 was used to finish RP11-224B10.

FEATURES	Location/Qualifiers
source	1. .183584

repeat_region	/rpt_family="MER1_type?"	6828. . 6948
repeat_region	7278. . 7391	
repeat_region	/rpt_family="MIR"	
repeat_region	7649. . 7786	
repeat_region	/rpt_family="Alu"	
repeat_region	8459. . 8507	
repeat_region	/rpt_family="(TG)n"	
repeat_region	9632. . 10223	
repeat_region	/rpt_family="L2"	
repeat_region	10360. . 11022	
repeat_region	/rpt_family="L2"	
repeat_region	11300. . 11326	
repeat_region	/rpt_family="(ATG)n"	
repeat_region	13389. . 13614	
repeat_region	/rpt_family="MIR"	
repeat_region	13960. . 13983	
repeat_region	/rpt_family="(TTG)n"	
repeat_region	14365. . 14552	
repeat_region	/rpt_family="MIR"	
repeat_region	14660. . 14764	
repeat_region	/rpt_family="L2"	
repeat_region	14918. . 15093	
repeat_region	/rpt_family="(TA)n"	
repeat_region	15283. . 15382	
repeat_region	/rpt_family="AT-rich"	
repeat_region	15603. . 15741	
repeat_region	/rpt_family="T-rich"	
repeat_region	15742. . 15913	
repeat_region	/rpt_family="(TTATA)n"	
repeat_region	16013. . 16195	
repeat_region	/rpt_family="(TA)n"	
repeat_region	16290. . 16452	
repeat_region	/rpt_family="(TA)n"	
repeat_region	16847. . 16940	
repeat_region	/rpt_family="MIR"	
repeat_region	20029. . 20424	
repeat_region	/rpt_family="MER1_type?"	
repeat_region	21038. . 21165	
repeat_region	/rpt_family="MIR"	
repeat_region	22953. . 23352	
repeat_region	/rpt_family="MaLR"	
repeat_region	23412. . 23500	
repeat_region	/rpt_family="MIR"	
repeat_region	23614. . 23701	
repeat_region	/rpt_family="ERV1"	
repeat_region	24258. . 24699	
repeat_region	/rpt_family="ERV1"	
repeat_region	25246. . 25416	
repeat_region	/rpt_family="L1"	
repeat_region	25505. . 25609	
repeat_region	/rpt_family="L1"	
repeat_region	27629. . 27760	
repeat_region	/rpt_family="MIR"	
repeat_region	29308. . 30562	
repeat_region	/rpt_family="L1"	
repeat_region	32823. . 33031	
repeat_region	/rpt_family="MIR"	
repeat_region	33530. . 33806	
repeat_region	/rpt_family="MaLR"	
repeat_region	34492. . 34795	
repeat_region	/rpt_family="Alu"	
repeat_region	36335. . 36465	
repeat_region	/rpt_family="L2"	
repeat_region	36484. . 36547	
repeat_region	/rpt_family="MIR"	
repeat_region	38315. . 38772	
repeat_region	/note="similar to EST AA705422 (NTD:92715340) zj90e10.s1"	
repeat_region	39308. . 39460	
repeat_region	/rpt_family="(CATAn"	
repeat_region	39488. . 39641	
repeat_region	/rpt_family="(TA)n"	
repeat_region	39962. . 40473	

```

repeat_region      /rpl_family="L1"
                    40474. .40773
                    /rpl_family="L1u"
                    40774. .41140
repeat_region      /rpl_family="L1"
                    41141. .42792
repeat_region      /rpl_family="L1"
                    42793. .42973
repeat_region      /rpl_family="L1"
                    43924. .44479
repeat_region      /rpl_family="L1"
                    44664. .44709
repeat_region      /rpl_family="L1"
                    47026. .47476
repeat_region      /rpl_family="ERV1"
                    48283. .48340
repeat_region      /rpl_family="MIR"
                    49380. .49607

```

Query Match 2.9%; Score 152.4; DB 9; Length 183584;
 Best Local Similarity 47.4%; Pred. No. 4e-14;
 Matches 682; Conservative 0; Mismatches 746; Indels 12; Gaps 7;

```

QY 33 ttattattccaaagtttaataatctcttgataacatccaactcttcacatgattga 92
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 16302 TATATAATATATATATATATATATATATATATATATATATATATATATATAT 16243

QY 93 ttgctgtaaaacccacagatttactacaataagggagcttgacttaaaatagctatg 152
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 16242 TTAATATATATATATATATATATATATATATATATATATATATATATATATAT 16183

QY 153 atgcgcgaaaaaatgtaatttagttataaattactaagaanaatttgcctgctg 212
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 16182 ATATATATATATATATATATATATATATATATATATATATATATATATATAT 16126

QY 213 tgttaagcataigtatgtgtaaaacttaaaaaaatgattgtttaacttcaaaatgta 272
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 16125 ATATATATATATATATATATATATATATATATATATATATATATATATATAT 16066

QY 273 ggaatcacacatacaatcgcagacataatcaaaacccgcatcatagacggatgtagaat 332
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 16065 TATATATATATATATATATATATATATATATATATATATATATATATATATAT 16006

QY 333 caaatagaagaatgtaatttttaaaatcgtatctccaatcaacttagaaga 392
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 16005 AATATATATATATATATATATATATATATATATATATATATATATATATATAT 15946

QY 393 taatgaattcttatagtgcacataaataatataatataatataatataatctt 452
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15945 TGAATATATATATATATATATATATATATATATATATATATATATATATATAT 15887

QY 453 gtatatatgcttgcaaaaaaattgccagtaaaaaacatgactgaatcaactaatagt 512
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15886 AAATATATATATATATATATATATATATATATATATATATATATATATATATAT 15827

QY 513 cggattgaatcaactataagtcgagatgaatcaattccatcatgcttctataactta 572
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15826 TGAATATATATATATATATATATATATATATATATATATATATATATATATAT 15767

QY 573 caaacgcgaaat-atagatattatagataccaaaaagtagattgtagataat-ata 630
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15766 TATATATATATATATATATATATATATATATATATATATATATATATATATAT 15707

QY 631 agatttgaatttcaatataatcagatctaaagtaactccctaattaatcaatgctcgt 690
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15706 AAATATATATATATATATATATATATATATATATATATATATATATATATATAT 15647

QY 691 tgaataagccaatgaatgcttgaaatttggaaagttatataatcggaactttttttt 750
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15646 ATATATATATATATATATATATATATATATATATATATATATATATATATATAT 15587

QY 751 ttgcttgcgcacaaacatttattattatacaaatcaacttaacttactacta 810
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 15586 ATATATATATATATATATATATATATATATATATATATATATATATATATAT 15527

QY 811 aatcatcacaatacttgaatcacacaacaaatcattcaatcttcttgatgattgaaga 870
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15526 AATATATATATATATATATATATATATATATATATATATATATATATATATAT 15467

QY 871 aacactattaccagttcaaaaatataatgaagcttgcttgtagaagaaagagacag 930
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15466 AATATATATATATATATATATATATATATATATATATATATATATATATATAT 15407

QY 931 ttgaattcttcttgcaaatataaattgacttttaatatataatgactattgaca 990
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15406 TATATATATATATATATATATATATATATATATATATATATATATATATATAT 15351

QY 991 tgatacgaattatcatcaccacaaacttccaaagtttataataatcaacttcaaga 1050
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15350 TATATATATATATATATATATATATATATATATATATATATATATATATATAT 15291

QY 1051 ctattagttcttcttaaaatatcttcaaaagtgatcacaagactcacacataatcaga 1110
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15290 TAAATATATATATATATATATATATATATATATATATATATATATATATATAT 15231

QY 1111 aaagttagaagttgacttcttcttgccaataaataatgacttaaaatagttgaaag 1170
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15230 TAAATATATATATATATATATATATATATATATATATATATATATATATATAT 15171

QY 1171 ccatggaacttgattatagaattgtaataatgtaacataaaatcccaagttataataa 1230
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15170 TAAATATATATATATATATATATATATATATATATATATATATATATATATAT 15111

QY 1231 tacatttccaagtatataatcagttcttcttaaaatatcttcaataaaacacatcaga 1290
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15110 TATATATATATATATATATATATATATATATATATATATATATATATATATAT 15052

QY 1291 talagaataaatttatgaataaacttacaactgtaaaacggaaatttgacaaaaaaa 1350
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15051 TATATATATATATATATATATATATATATATATATATATATATATATATATAT 14992

QY 1351 aaaaatgaatgaagaatgaagacaaaaataatcacacagagatctatgcaaaaaat 1410
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14991 AATATATATATATATATATATATATATATATATATATATATATATATATATAT 14933

QY 1411 atatgaataacaaataacatattgatalactttaaaataaataaacaagaaaaat 1470
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14932 ATAAATATATATATATATATATATATATATATATATATATATATATATATATAT 14873

```

Search completed: May 13, 2002, 07:58:43
 Job time: 19492 sec


```

Db 625 -----agagccatgcggatattcaacccctgtgtagtttcttctgtcgaac 677
Oy 661 agtagcctcatcgaataatctatgaagtaagttgtttgatcgaacggattctg 720
Db 678 agtagcctcatcgaataatctatgaagtaagttgtttgatcgaacggattctg 737
Oy 721 tttagtcgaagtcgtcgaagatcttcgacgtttctatgtgttatcccttcgac 780
Db 738 tttagtcgaagtcgtcgaagatcttcgacgtttctatgtgttatcccttcgac 797
Oy 781 ttcccttggcgttccttaccgtttggaattgttactcagaataacatcagaacct 840
Db 798 ttcccttggcgttccttaccgtttggaattgttactcagaataacatcagaacct 857
Oy 841 gttgtcatcttctcatatattatcacacgacgaaggtttgtatccagttagtc 900
Db 858 gttgtcatcttctcatatattatcacacgacgaaggtttgtatccagttagtc 917
Oy 901 accctaaggctgcatctgtcttttatacagtgatcacttgcctcctcactgcat 960
Db 918 accctaaggctgcatctgtcttttatacagtgatcacttgcctcctcactgcat 977
Oy 961 gttgtcgaagttgttcttctatctcatctactatgataatagaatccctagcaat 1020
Db 978 gttgtcgaagttgttcttctatctcatctactatgataatagaatccctagcaat 1037
Oy 1021 gtagctgaataagccaatcctctgaagtcctcactacgttgaagcttgaagcttgcacat 1080
Db 1038 gtagctgaataagccaatcctctgaagtcctcactacgttgaagcttgaagcttgcacat 1097
Oy 1081 ttcatgctgcgtccacacgttgttatcagaacgttatccagcttctgcatacag 1140
Db 1098 ttcatgctgcgtccacacgttgttatcagaacgttatccagcttctgcatacag 1157
Oy 1141 aaggttgggtgctgcgaatttgcgaacgtgtcatcaccgagatcattgagatt 1200
Db 1158 aaggttgggtgctgcgaatttgcgaacgtgtcatcaccgagatcattgagatt 1217
Oy 1201 ataagaacaataataataacctaattgtcgaagaaactcaacgactcttgaaggcgat 1260
Db 1218 ataagaacaataataataacctaattgtcgaagaaactcaacgactcttgaaggcgat 1277
Oy 1261 ctctcatactgcatgaagaaggtgtgaagcttcaagttccaatttatgtgtgctc 1320
Db 1278 ctctcatactgcatgaagaaggtgtgaagcttcaagttccaatttatgtgtgctc 1337
Oy 1321 tgcattgtctactgctcttccacgttgttgaacatattgagagagcttctgctc 1380
Db 1338 tgcattgtctactgctcttccacgttgttgaacatattgagagagcttctgctc 1397
Oy 1381 ggggagctggaattctacaagaattgtggaatgcaaaaagtgtggagattactgaga 1440
Db 1398 ggggagctggaattctacaagaattgtggaatgcaaaaagtgtggagattactgaga 1457
Oy 1441 atgtggaatctgctgttcatataatgtgagttgcacatatatactccgtgtgcgc 1500
Db 1458 atgtggaatctgctgttcatataatgtgagttgcacatatatactccgtgtgcgc 1517
Oy 1501 agcaagaatacacaagaacatgcgcattatacattgtcttccctagtcctgcagttcat 1560
Db 1518 agcaagaatacacaagaacatgcgcattatacattgtcttccctagtcctgcagttcat 1577
Oy 1561 gagctatgcatcgagttcctgtcgtctcttcaagctatggggtttctgtgattatg 1620
Db 1578 gagctatgcatcgagttcctgtcgtctcttcaagctatggggtttctgtgattatg 1637
Oy 1621 tttagagttgcttgccttcatcacaacatctatcagaagaagttgtgctcaagctg 1680
Db 1638 tttagagttgcttgccttcatcacaacatctatcagaagaagttgtgctcaagctg 1697
Oy 1681 ggggaagtgtcttctgttcatcttctgtgaatttgcgaacccgagtgtgtctctt 1740

```

```

Db 1698 ggggaagtgtcttctgttcatcttctgtgaatttgcgaacccgagtgtgtctct 1757
Oy 1741 tatlaccagacctgtagtaacccaagaagatcgaatgcatgaatacaacttcaaaaat 1800
Db 1758 tatlaccagacctgtagtaacccaagaagatcgaatgcatgaatacaacttcaaaaat 1817
Oy 1801 gacttcttcaaacatctatgctcgttggatctccgttgaatgtgtgtgtctgac 1860
Db 1818 gacttcttcaaacatctatgctcgttggatctccgttgaatgtgtgtgtctgac 1877
Oy 1861 gctaaacgacaataagtgatataaccattgaagaagaagaataatagaattgttga 1920
Db 1878 gctaaacgacaataagtgatataaccattgaagaagaagaataatagaattgttga 1937
Oy 1921 tc 1922
Db 1938 tc 1939

RESULT 6
AAS01106
ID AAS01106 standard; cDNA, 1942 BP.
AC AAS01106;
DX 31-MAY-2001 (first entry)
DE Arabidopsis thaliana sterol acyltransferase ACAT EST sequence.
KW Lecithin:cholesterol acyltransferase-like; LCAT; sterol acyltransferase;
KW acyl CoA:cholesterol acyltransferase-like; ACAT; plant oil; cooking oil;
KW nutritional supplement; dairy product; food product; salad dressing;
KW expressed sequence tag; EST; ss.
OS Arabidopsis thaliana.
FH Key Location/Qualifiers
FT CDS 237..1799
FT /tag= a
FT /product= "ACAT"
PN MO200116308-A2.
PD 08-MAR-2001.
XX 30-AUG-2000; 2000MO-US23863.
XX 30-AUG-1999; 99US-0152493.
XX (MONS ) MONSANTO CO.
XX Lasser M, Van Benennaam A;
XX MPI; 2001-169010/17.
XX P-PSDB; AAU00462.
XX
XX New isolated nucleic acid encoding plant lecithin:cholesterol
XX acyltransferase-like or acyl (coenzyme A) CoA:cholesterol
XX acyltransferase-like polypeptides, for modifying the sterol content and
XX oil production of plants
XX
XX Claim 4; Page 105-106; 127pp; English.
XX
XX The present sequence encodes for Arabidopsis thaliana
XX acyl CoA:cholesterol acyltransferase-like (ACAT). Several novel
XX polynucleotides encoding the plant sterol acyltransferases LCAT
XX (lecithin:cholesterol acyltransferase-like; AAS001081-AAS01104, AAS01341)
XX and ACAT (AAS01311-AAS01319) are described. A yeast LCAT related open
XX reading frame, LMO1 gene sequence (AAS01342), and a rat ACAT (AAS01105)
XX cDNA sequence are also described. The polynucleotides encoding LCAT
XX or ACAT are used to produce LCAT or ACAT polypeptides. They can also
XX be used in a recombinant construct to transform a host cell (preferably
XX of a plant) or a plant. The recombinant construct is used to increase or

```


Oy		1921 tc	1922
I/I			
Dd		1938 tc	1939
RESULT 7 AAAA8939 standard; cDNA; 1942 BP.			
ID		AAAA8939;	
XX		AAA48933;	
AC		66-DEC-2000	(first entry)
XX			
DT		Soybean diacylglycerol acyltransferase cDNA #1.	
XX			
DE		Diacylglycerol acyltransferase; corn; rice; soybean; wheat;	
KM		triacylglycerol; herbicide; EC2.3.1.20; SS.	
XX		Glycine max.	
OS			
XX			
FH		Location/Qualifiers	
FT		/tag= a	
CDS		/product= Diacylglycerol_acyltransferase	
PN		M0200032756-A2.	
PX		08-JUN-2000.	
PD		01-DEC-1999;	99WO-US28354.
PF		02-DEC-1998;	98US-0110602.
PR		31-MAR-1999;	99US-0127111.
PA		(DUBO) DU PONT DE NEMOURS & CO E I.	
XX			
PI		Cahoon EB, Kinney AJ, Cahoon RE;	
DR		WPI: 2000-412308/35.	
XX		P-PSDB: MMY94519.	
PT		Polynucleotides encoding diacylglycerol acetyltransferase, useful for	
PS		synthesis of triacylglycerols and increasing the level of oils in plant	
XX		seeds -	
Claim 4;		Page 51; 62pp; English.	
CC		In the present invention, cDNA libraries from Arabidopsis, corn, rice,	
CC		soybean and wheat were screened for sequences with homology to a	
CC		putative acyl CoA cholesterol acyltransferase related gene from	
CC		Arabidopsis thaliana and diacylglycerol acyltransferases from Homo	
CC		sapiens and Mus musculus. The cDNA clones identified from this process	
CC		were used to form complete diacylglycerol acyltransferase cDNA	
CC		sequences. The present sequence is soybean diacylglycerol	
CC		acyltransferase cDNA from clone srl.pk0098.a8. Diacylglycerol	
CC		acyltransferases are involved in the synthesis of triacylglycerols.	
CC		Alteration of the expression of the diacylglycerol acyltransferase	
CC		cDNA can be useful for increasing the level of oils in plant seeds.	
CC		Inhibitors of diacylglycerol acyltransferase may be useful as	
herbicides.			
SO		Sequence 1942 BP; 517 A; 425 C; 377 G; 623 T; 0 other;	
Query Match		34.4%; Score 683; DB 21; Length 1942;	
Best Local Similarity		74.4%; Pred. No. 7e-187;	
Matches 899;		Conservative 0; Mismatches 285; Indels 24; Gaps 2	
db		608 agagccatgcggattatccaacctgtgtagtactcttatcttcgtlaaacagtac 667	
OY		345 agagctacgaggcccttcaacccttgatagtactctttgtcgctgaatgccgac 404	
OY		668 tcatcatcgaaatcttataagaatgatgttgtttgattgataagaagattctgtttaagt 727	

Db	405	tcatctcttagaatttaataagaaatggttggttgatgacaatctgcttggttagtc	464
Qy	728	caagatcgcgcggaaggttgcgcgctttccatggtgtgtgatacccttgcatacttcct	787
Db	465	caaaagcattgaaagagctggccctcttccatgtgtgttccttccctgtgtgatttcct	524
Qy	788	ttgcgtcccttcaacgtctggaataatgtgacttcagaaatacatcaagaaccgttgtca	847
Db	525	tcgtgccttatagtagggagaagttggacaacgaagtgatataccggaaccagttgttg	584
Qy	848	tcttcttcaatatattatcacaccatgacagaagttttgtatccagttttacgcacctaa	907
Db	585	tgtactatcaataatcaattcaactcaacttcgcttcttccatccagtttagttatctoca	644
Qy	908	gggtgatctcctcttttattatcaggtgtgcacttttgatgcctccacttgcattgtgtgc	967
Db	645	gggtgatctcctcttttattatcaggtgtgcacggttaagtcggtttctctgtgtgtatgt	704
Qy	968	taaaagtgttctctatgtccataactagctatgacataagaatccctgaccacatgcagtcg	1021
Db	705	taaaatgtgtcctatgtcacatacaaacctatgtatagagacacttaccaaattagttg	764
Qy	1028	atgaag-----ccatccctgaagctcctactaagttagctgtga	1066
Db	765	aaaagggaaagacactgcctgcatactctgacaatgacatactctccaaagtattcca	824
Qy	1067	agaagcttggcatatttcaatgtgtcgtcccaacatgtgttialcagccaagtatccacgtt	1120
Db	825	agaagcttggcatatttctcgtgtgcccctacataattgtttacacgaacgaatccctcgca	884
Qy	1127	cgtcagtgatacgggaaggttgggttggtcgtgcataattggaaaactcgttcatataccg	1186
Db	885	caacctaatctcgaaagggttgcgttgttctgcgcaactgttgcacgctgataatattcacag	944
Qy	1187	gattcatggaatttaataagacaacataataacctatgttcaggaactcaaaagctt	1246
Db	945	gggttatggatttataatagaccataataattatcccatagacaaaatttcaacgctc	1004
Qy	1247	ctttgaagcgcgactcttctataatgtctatgaaagatgtgtgaagcttcagttccaant	1306
Db	1005	ctctcaaaaggaaacctcttcttaccgcacacggagagttctcgaaagcttctcgttccaant	1066
Qy	1307	tataatgtgtgcctcgcgaatgttctcaacgctctcttccacacttgggttaaacatattggag	1366
Db	1065	latatgtgtgctctcgtgaatgttccatgtcttlttccacttgggttaaatccctggcgaag	1124
Qy	1367	agctctctcgtctcggggatcgtgaattctacaaaagatttgggtgaatgtgaaaagtgtgg	1426
Db	1125	agctctctcgaatttggtagacgtgaaattctacaagatttgggtgaatgtcaaaactgtcg	1184
Qy	1427	gagatttactggaagaigtggaataatgcgtttcaataaagatgtgtcgcacataatact	1486
Db	1185	aagattatctggaaggaatgtggaataatgtcgtttcaacaatggatgtccgcacactatt	1244
Qy	1487	tcccggtgttcggaagaaatagatacacaagaacatcgcacatatacttgccttcacagct	1546
Db	1245	ttccatgtttaaggcacggtctacaaagcgtcgtccttttaattgtccttccgtgttt	1304
Qy	1547	cttgcagcttcttcatgagctatgtcatcgcagttccctgttcgtctcttcaagctatggctt	1606
Db	1305	ctgcttattccacatgacgtgtgatctgtcttccctgtccacacatactcaagtgttggctt	1364
Qy	1607	tctcttggatatagtttccagtgaccttgggttccatcaacaacatacctaagaagaagt	1666
Db	1365	tcgttggatatagtttccagtgaccttgggttccatcaacaacatacctaagaagaagt	1424
Qy	1667	tttgg-----ctcaacggttggggacaatgactctcgttgcatacttctgcatttctggacaac	1723
Db	1425	tcgaagaactcaatggttggaaataatgatttttggttccatattcaagtacttggtaaac	1484
Qy	1724	cgatgtgtgtccttcttataccacgacctgtgaacacgaaagaatcgtatgtcatgaa	1783

```

Db 1485 ctatgctgtactgtactactgtactgtatgaatagaagaagcaactgactgaa 1544
Oy 1784 acaactg 1790
    |||
Db 1545 gctacg 1551

RESULT 8
AAAS1483
ID AAAS1483 standard; DNA: 5193 BP.
XX
AC AAAS1483;
XX
XX 09-OCT-2000 (first entry)
XX
DE A. thaliana diacylglycerol acyltransferase genomic DNA.
XX
XX DGAT; diacylglycerol acyltransferase; seed oil; fatty acid synthesis;
XX size; weight; carbon flux; ds.
XX
XX Arabidopsis thaliana.
XX
XX Key Location/Qualifiers
FH 1761..4780
FT /*tag= a
FT 1761..2150
FT exon
FT /*tag= b
FT /partial
FT 2151..2234
FT /*tag= c
FT 2223..2369
FT /*tag= d
FT /note= "region duplicated in insertion mutant AS11
      (See AAAS1485)"
FT
FT exon
FT /*tag= e
FT 2316..2406
FT /*tag= f
FT 2407..2480
FT exon
FT /*tag= g
FT 2481..2582
FT /*tag= h
FT 2583..2661
FT /*tag= i
FT 2662..2758
FT intron
FT /*tag= j
FT 2759..2823
FT exon
FT /*tag= k
FT 2824..2927
FT intron
FT /*tag= l
FT 2928..3028
FT exon
FT /*tag= m
FT 3029..3108
FT intron
FT /*tag= n
FT 3109..3186
FT exon
FT /*tag= o
FT 3187..3264
FT intron
FT /*tag= p
FT 3265..3366
FT exon
FT /*tag= q
FT 3367..3466
FT intron
FT /*tag= r
FT 3467..3603
FT exon
FT /*tag= s
FT 3604..3747
FT intron
FT /*tag= t
FT 3748..3829
FT exon
FT /*tag= u
FT 3830..3933
FT intron
FT /*tag= v
FT 3934..3954
FT exon
FT /*tag= w
FT 3955..4038
FT intron

```

```

FT /*tag= x
FT 4039..4101
FT /*tag= y
FT 4102..4184
FT intron
FT /*tag= z
FT 4185..4232
FT exon
FT /*tag= aa
FT 4233..4355
FT intron
FT /*tag= ab
FT 4356..4418
FT exon
FT /*tag= ac
FT 4419..4499
FT intron
FT /*tag= ad
FT 4500..4550
FT exon
FT /*tag= ae
FT 4551..4675
FT intron
FT /*tag= af
FT 4676..4780
FT exon
FT /*tag= ag
FT /partial

```

W0200036114-A1.

XX 22-JUN-2000.

XX 16-DEC-1999; 99WO-CA01202.

XX 17-DEC-1998; 98US-0112812.

XX (CANADA) NAT RES COUNCIL CANADA.

XX Zou J, Taylor DC, Wei Y, Jako CC;

XX WPI: 2000-431592/37.

XX P-PSDB: AAY96853.

XX New DNA encoding diacylglycerol acyltransferase from Arabidopsis
 PT thaliana for transforming plants and regulating seed oil content, fatty
 PT acid synthesis and seed oil acyl composition in commercial and crop
 PT plants

XX Claim 2; Page 67-68; 91pp; English.

XX This DNA encodes Arabidopsis thaliana diacylglycerol acyltransferase
 CC (DGAT). The DGAT and an insertion mutant (AS11) are useful for regulating
 CC seed oil content, the ratio of diacylglycerol/triacylglycerol proportions
 CC in seed oil, fatty acid synthesis, seed oil acyl composition, seed
 CC size/weight and carbon flux into other seed components in commercial and
 CC crop plants. The natural formation of triacylglycerols can be modified to
 CC increase the yield in commercial plant oils or modify their composition
 CC to achieve specific commercial improvements of plants and plant products.
 CC
 XX Sequence 5193 BP; 1582 A; 894 C; 903 G; 1814 T; 0 other;

Query Match

29.9%; Score 592.6; DB 21; Length 5193;
 Best Local Similarity 91.2%; Pred. NO. 1; 7e-160;
 Matches 641; Conservative 0; Mismatches 59; Indels 3; Gaps 1;

```

Oy 1 attcttagctctctcctcaatccgcctcttccctccatagatctgttccctt 60
    |||||||
Db 1623 attcttagctctctcctcaatccgcctcttccctccatagatctgttccctt 1682
Oy 61 tcaattcttcgcagctcttcgattctctctgagcgctcttctccgaagctgtt 120
    |||||||
Db 1683 tcaattcttcgcagctcttcgattctctctgagcgctcttctccgaagctgtt 1742
Oy 121 cgtcaacgcttctcgaatcttgagattctctgctgagcggtactacggtgagag 180
    |||||||
Db 1743 cgtcaacgcttctcgaatcttgagattctctgctgagcggtactacggtgagag 1802
Oy 181 aacggtgagcgaggtctgctgattgagatctgctgagcggaatcgagatcgat 240
    |||||||

```


PS Disclosure: Page 79-81: 91pp; English.

XX This DNA encodes Arabidopsis thaliana ecotype Columbia mutant AS11
CC diacylglycerol acyltransferase (DGAT). The AS11 mutant TAG1 allele has a
CC 147 bp insertion located at the central region of intron 2. The insertion
CC is a duplication of a segment that is composed of 12 bp from the 3' end
CC of intron 1, the entire sequence of exon 2 (81 bp) and 54 bp from the
CC 5' end of intron 2. The DGAT and the insertion mutant (AS11) are useful
CC for regulating seed oil content, the ratio of diacylglycerol to
CC triacylglycerol proportions in seed oil, fatty acid synthesis, seed oil
CC acyl composition, seed size/weight and carbon flux into other seed
CC components in commercial and crop plants. The natural formation of
CC triacylglycerols can be modified to increase the yield in commercial
CC plant oils or modify their composition to achieve specific commercial
CC improvements of plants and plant products.

XX Sequence 5339 BP; 1620 A; 919 C; 930 G; 1870 T; 0 other;

Query Match 29.9%; Score 592.6; DB 21: Length 5339;

Best Local Similarity 91.2%; Pred. No. 1.7e-160;

Matches 641; Conservative 0; Mismatches 59; Indels 3; Gaps 1;

QY 1 attcttagctctctccctcaatccgcctctccctccattagattcgttccctct 60
DB 1623 attcttagctctctccctcaatccgcctctccctccattagattcgttccctct 1682
QY 61 tcaattctctctgct 120
DB 1683 tcaattctctctgct 1742
QY 121 cgtcaaacgctcttctgcaaatgagcatttgagctctgctgctgcttactacggagagag 180
DB 1743 cgtcaaacgctcttctgcaaatgagcatttgagctctgctgctgcttactacggagag 1802
QY 181 aacgctgagcagagagctctgctgctgctgctgctgctgctgctgctgctgctgct 240
DB 1803 aacgctgagcagagagctctgctgctgctgctgctgctgctgctgctgctgctgct 1862
QY 241 tcttctaacgagct 300
DB 1863 tcttctaacgagct 1922
QY 301 gctcccgccgagctttagagagctcagattgattcgttctgtaacagatgagctcagagag 360
DB 1923 gctcccgccgagctttagagagctcagattgattcgttctgtaacagatgagctcagagag 1982
QY 361 gccatttggccgagataataacagctgctgctgctgctgctgctgctgctgctgctgct 420
DB 1983 gccatttggccgagataataacagctgctgctgctgctgctgctgctgctgctgctgct 2042
QY 421 ggagaag 480
DB 2043 ggagaag 2102
QY 481 cggagagcag 540
DB 2103 cggagagcag 2162
QY 541 ttattcaacctctgtgtagtagttcttattgctgttaaacagtagactatcctgaaat 600
DB 2163 agaaatcttcgaattgt 2222
QY 601 ctt---atgaagagcagatcgagatattcaacctctgtgtgtgtgtgtgtgtgtgtgt 657
DB 2223 ttgtgattgag 2282
QY 658 aacagtagactatcctgaaataacttaagaagtagtgtgtgtgtgtgtgtgtgtgtgt 700
DB 2283 aacagtagactatcctgaaataacttaagaagtagtgtgtgtgtgtgtgtgtgtgtgt 722
2283 aacagtagactatcctgaaataacttaagaagtagtgtgtgtgtgtgtgtgtgtgtgt 787
2283 aacagtagactatcctgaaataacttaagaagtagtgtgtgtgtgtgtgtgtgtgtgt 852

RESULT 10

AAA48942
ID AAA48942 standard; cDNA; 1975 BP.

XX AAA48942;

XX 06-DEC-2000 (first entry)

XX Wheat diacylglycerol acyltransferase cDNA #2.

XX Diacylglycerol acyltransferase; corn; rice; soybean; wheat;

XX triacylglycerol; herbicide; EC2.3.1.20; ss.

XX Triticum aestivum.

XX Key location/Qualifiers

XX CDS 107..1633

XX /tag= a

XX /product= Diacylglycerol_transferase

XX WO200032756-A2.

XX 08-JUN-2000.

XX 01-DEC-1999; 99WO-US28354.

XX 02-DEC-1998; 98US-0110602.

XX 31-MAR-1999; 99US-0127111.

XX (DUFO) DU POINT DE MEMOIRS & CO E. I.

XX Cahoon EB, Kinney AJ, Cahoon RE;

XX WPI: 2000-412308/35.

XX P-PSDB: AAY94522.

XX Polynucleotides encoding diacylglycerol acetyltransferase, useful for

XX synthesis of triacylglycerols and increasing the level of oils in plant

XX seeds

XX Claim 4: Page 56; 62pp; English.

XX In the present invention, cDNA libraries from Arabidopsis, corn, rice,

XX soybean and wheat were screened for sequences with homology to a

XX putative acyl CoA cholesterol acyltransferase related gene from

XX Arabidopsis thaliana and diacylglycerol acyltransferases from Homo

XX sapiens and Mus musculus. The cDNA clones identified from this process

XX were used to form complete diacylglycerol acyltransferase cDNA

XX sequences. The present sequence is wheat diacylglycerol

XX acyltransferase cDNA from clone wrl.pk0119.b6.fis. Diacylglycerol

XX acyltransferases are involved in the synthesis of triacylglycerols.

XX Alteration of the expression of the diacylglycerol acyltransferase

XX DNA can be useful for increasing the level of oils in plant seeds.

XX Inhibitors of diacylglycerol acyltransferase may be useful as

XX herbicides.

Sequence 1975 BP; 463 A; 497 C; 461 G; 550 T; 4 other;

Query Match 27.4%; Score 543.6; DB 21: Length 1975;

Best Local Similarity 66.5%; Pred. No. 1.3e-146;

Matches 821; Conservative 0; Mismatches 389; Indels 24; Gaps 2;

QY 608 agagc 667
DB 432 agagc 491
QY 668 tcatcatcgaataactatgaagtagtctgtgtgatacgaagcagcagcagcagcagcagcagcagc 727
DB 492 tcatcatcgaagaaactatgaagtagtctgtgtgatacgaagcagcagcagcagcagcagcagc 551
QY 728 caagatcgtcagagagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 787
DB 552 caagatcgtcagagagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 611

```

QY 768 tggctgcctttagcgttgagaattggtactcagaataatatacatagaaacctgtgtca 847
Db 612 ttgctgcctttagcgttgagaattggtactcagaataatatacatagaaacctgtgtca 671
QY 848 tcttctcattatattatcaccatgacagagtttctgtatccagtttgaacacctaa 907
Db 672 tcttctcattatattatcaccatgacagagtttctgtatccagtttgaacacctaa 731
QY 908 ggtgtgacttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 967
Db 732 agtgtgacttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 791
QY 968 taaagttggttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1027
Db 792 tgaagctgttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 851
QY 1028 ataagcccaatcctggaagcttccctactagtt-----agcttga 1066
Db 852 aaaaagggtgctacacacagcagttctatcgatgaggaanaaatgaagcccaactatca 911
QY 1067 agagctgtgacatattatctgtcgcctccacatgtgtatccagcaagtataccagct 1126
Db 912 acagctgtgtgtattctatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 971
QY 1127 ctgcattatagcgaagaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1186
Db 972 cagcattatagcgaagaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1031
QY 1187 gattcattggaattataatagaacaaatataatccatctatgtcagaaactcaaacac 1246
Db 1032 gctttagtggcttcaatagcaataatcattatccaatgtgtgtgtgtgtgtgtgtgt 1091
QY 1247 ctttgaagagcgatctcttcttcttcttcttcttcttcttcttcttcttcttctt 1306
Db 1092 catggaacggaaattctctgtgagcattgtgagagcttcttgaactctgaagccaaac 1151
QY 1307 tatagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1366
Db 1152 tatatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1211
QY 1367 agcttctgtcttgcggaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1426
Db 1212 aactctcgcgttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1271
QY 1427 ggaattactggaagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1486
Db 1272 aagaatgactggaagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1331
QY 1487 tccgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1546
Db 1332 ttcctatgataaaggaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1391
QY 1547 ctgcagcttctatgagatgacatgacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1606
Db 1392 cagctgtatctatgagatgacatgacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1451
QY 1607 tttctgagatattgttttaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1666
Db 1452 tttctgagatattgttttaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1511
QY 1667 t---tggctcaacgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1723
Db 1512 tcaagaatacaaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1571
QY 1724 cgaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1783
Db 1572 caatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1631
QY 1784 acaactgtcaaaaaaagacttcttcaaacatc 1817
Db 1632 agttctgttttaggaagtgcactataacagagac 1665

```

RESULT 11

AAA48938

ID AAA48938 standard; cDNA; 1587 BP.

XX AAA48938;

DT 06-DEC-2000 (first entry)

DE Rice diacylglycerol acyltransferase cDNA #2.

KW Diacylglycerol acyltransferase; corn; rice; soybean; wheat;
KM triacylglycerol; herbicide; EC2.3.1.20; ss.

XX Oryza sativa.

FH Key Location/Qualifiers

FT CDS 15..1310

FT /tag=a

FT /partial

FT /note="The CDS of this sequence only encodes amino acids 70 to 500 of the protein in AA94518"

PN WO200032756-A2.

PD 08-JUN-2000.

PF 01-DEC-1999; 99WO-US28354.

PR 02-DEC-1998; 98US-0110602.

PR 31-MAR-1999; 99US-0127111.

PA (DUPO) DU PONT DE NEMOURS & CO E. I.

PI Cahoon EB, Kinney AJ, Cahoon RE;

DR WPI: 2000-412308/35.

DR P-PSDB: AA94518.

PT Polynucleotides encoding diacylglycerol acetyltransferase, useful for

PT synthesis of triacylglycerols and increasing the level of oils in plant

PT seeds

PS Claim 4; Page 48-49; 62pp; English.

XX In the present invention, cDNA libraries from Arabidopsis, corn, rice,

XX soybean and wheat were screened for sequences with homology to a

XX putative acyl CoA cholesterol acyltransferase related gene from

XX Arabidopsis thaliana and diacylglycerol acyltransferases from Homo

XX sapiens and Mus musculus. The cDNA clones identified from this process

XX were used to form complete diacylglycerol acyltransferase cDNA

XX sequences. The present sequence is rice diacylglycerol acyltransferase

XX cDNA from clone R1824.pK0034.d8.f15. Diacylglycerol acyltransferase are

XX involved in the synthesis of triacylglycerols. Alteration of the

XX expression of the diacylglycerol acyltransferase DNA can be useful for

XX increasing the level of oils in plant seeds. Inhibitors of

XX diacylglycerol acyltransferase may be useful as herbicides.

SQ Sequence 1587 BP; 411 A; 323 C; 334 G; 519 T; 0 other;

Query Match 25.2%; Score 500.8; DB 21; Length 1587;

Best Local Similarity 65.4%; Pred. No. 3.3e-134;

Matches 777; Conservative 0; Mismatches 387; Indels 24; Gaps 2;

QY 602 ttatgaagacatccgcatattcaacctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 661

Db 106 tcaagcagagtcacagagccttccaacctatgatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 165

QY 662 gttagactatcatgaaatcttatgaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 721

Db 166 gcaagctattatcagaaacttaagtaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 225


```

DB 181 ctgcgattcttaagtgtgattctgcagttatatacaggcttctgtgtatgttatttgc 240
OY 955 tgcattgtgtgctaaagtgttcttctatgctacatacagatgataagatcccta 1014
DB 241 tgcattgtgtgctaaagtgttcttctatgctacatacagatgataagatcccta 300
OY 1015 gccaatgcagctgataagggccaatccctgaagctctcctactaagttagcttga 1067
DB 301 atcacaagcgcaagaaggttgataatgaactgacgcgctgctgataatattaca 360
OY 1068 -----gagcttgcatattcattcagctgcgcgcacacatgtgtatcagcca 1113
DB 361 gctccaactctgtgagagctcaacatacctcagatgagctgcacactcgttatagcca 420
OY 1114 agttatccagcttgcagatgatacgaagggtgtgtgtcgtcgtcaattgcaaaactg 1173
DB 421 agttatccctgcaaacactatgttagaaaggtgtgtcgtcgtcgaagtattctatac 480
OY 1174 gtcaattcaccgagatcaggtattataatagaacaataataatcctattgtcagg 1233
DB 481 ttgatatttactgtgtctccaagatcattatgagcaataataatcctattgtgtg 540
OY 1234 aactcaagacatccttgaaggcagctctctatagctatgataagaggtgtgaagct 1293
DB 541 aactccaacatcattgtgtgagagattactgaagctgtgtagagactgttttgaagctc 600
OY 1294 tcaattccaattatattgtgtgtcgtcagatgttctactgcttctccactgtgtta 1353
DB 601 tcaatcaaaatgtctaccgtgtgtgtgttattgttactgttctccactgtgtta 660
OY 1354 aacatttggcagagcttctgtcgtcggagctgagatcttcaaaagattgtgtgaat 1413
DB 661 aacatttgcgtgagcttctgtcgtcggagctgagatcttcaaaagattgtgtgaat 720
OY 1414 gcaaaaagtgtggaagattactgaggaatgtggaatagatagctgttataaagtgtgt 1473
DB 721 gcaaaagcaattgtagtactgaggaatgtggaatagatagctgttataaagtgtgt 780
OY 1474 cgaatatataactcccgctgtgtgcgcaagaatacagaagacatcccatattatc 1533
DB 781 cgtatataatttcttctgtcagtgcaaaatggtatatacagaagaattgtgtttata 840
OY 1534 gcttctagctctcgtcagcttctcagctta 1566
DB 841 tctgtcttctgtctcgtcgttactcattagta 873

```

RESULT 14 AAA48936

ID AAA48936 standard; cDNA; 901 BP.

XX AAA48936;

XX 06-DEC-2000 (first entry)

XX corn diacylglycerol acyltransferase cDNA #4.

XX Diacylglycerol acyltransferase; corn; rice; soybean; wheat;

XX triacylglycerol; herbicide; EC2.3.1.20, ss.

XX Zea mays.

XX OS

XX Key

XX CDS

FT 1.858 location/Qualifiers
FT //tag= a
FT //partial
FT //transl_except= (pos:442..444,aa:Xaa)
FT //transl_except= (pos:490..492,aa:Xaa)
FT //transl_except= (pos:587..589,aa:Xaa)
FT //transl_except= (pos:805..807,aa:Xaa)
FT //transl_except= (pos:820..822,aa:Xaa)
FT //note= "Xaa= unknown"

FT /product= Partial_diacylglycerol_acyltransferase

XX W0200032756-A2.

XX 08-JUN-2000.

XX 01-DEC-1999; 99MO-US28354.

XX 02-DEC-1998; 98US-0110602.

XX 31-MAR-1999; 99US-0127111.

XX (DUPO) DU PONT DE NEMOURS & CO E. I.

XX Cahoon EB, Kinney AJ, Cahoon RE;

XX WPI: 2000-412308/35.

XX P-PSDB: AAY94516.

XX Polynucleotides encoding diacylglycerol acetyltransferase, useful for
XX synthesis of triacylglycerols and increasing the level of oils in plant
XX seeds

XX Claim 18; Page 46; 62pp: English.

XX In the present invention, cDNA libraries from Arabidopsis, corn, rice,
XX soybean and wheat were screened for sequences with homology to a
XX putative acyl CoA cholesterol acyltransferase related gene from
XX Arabidopsis thaliana and diacylglycerol acyltransferases from Homo
XX sapiens and Mus musculus. The cDNA clones identified from this process
XX were used to form complete diacylglycerol acyltransferase cDNA
XX sequences. The present sequence is corn p0042.csp49t, p0122.cbam57t and
XX cDNA from a contig of clones p0042.csp49t, p0122.cbam57t and
XX p0125.czaa61rb. Diacylglycerol acyltransferases are involved in the
XX synthesis of triacylglycerols. Alteration of the expression of the
XX diacylglycerol acyltransferase DNA can be useful for increasing the
XX level of oils in plant seeds. Inhibitors of diacylglycerol
XX acyltransferase may be useful as herbicides.

XX Sequence 901 BP; 256 A; 172 C; 193 G; 276 T; 4 other;

XX Query Match 17.9%; Score 355.2; DB 21; Length 901;

XX Best Local Similarity 67.1%; Pred. No. 3.2e-92;

XX Matches 501; Conservative 0; Mismatches 246; Indels 0; Gaps 0;

XX 1028 ataaggccaatcctgaagcttctactactagcttgaagagcttgcacattcagtg 1087

XX 104 attatgtgactcctggaatatacgaatcgaactttaaagcttagtacttcatgt 163

XX 1088 tgcgtccacattgtgttatcagccaagtatccagcttgcagtgatagcgaagggt 1147

XX 164 tggccccaacacttgttaccagcaactatccccaactacatgatatagaagggtt 223

XX 1148 ggggtgctgcgaattggaacactgtcatatacccgatcattgagattatagaag 1207

XX 224 ggggtgccaagcaaccataaagtgtgttttaacagcttggcttcataatgtg 283

XX 1208 aacaalataaactctatgttcaggaactcaagacatccttgaagaagcgtctcat 1267

XX 284 agcaatataaaccgaatgtgaagaattccaaacatccactgaaaggaatttga 343

XX 1268 atgctatgaagaagagtgtaagcttgcagttccaaattacatagtgtgctcagatg 1327

XX 344 atgctatgaagaagagtgtaagcttgcagttccaaattacatagtgtgctcagatg 403

XX 1328 tctactgcttctccacttgtttaaacaattggcagagcttctcgttcggagatc 1387

XX 404 tctattgcttcttctattatgtgtgaacttgaacttgaactcctcgttctgagacc 463

XX 1388 gtaaatctcaagaagatgtgtgaatgaacaaagtgtgtgagatcttggaagtgtga 1447

XX 464 gtaaatctcaagaagatgtgtgaatgaacaaagtgtgtgagatcttggaagtgtga 523

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 07:58:43 ; Search time 8260.66 Seconds
(without alignments)
5028.554 Million cell updates/sec

Title: US-09-623-514A-23

Perfect score: 1985
Sequence: 1 attcttagctctctctc.....ttatgtgttaagcgccgc 1985

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pt:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_hugo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	1813	91.3	1904	8	ATH238008	AJ238008 Arabidops
2	1805	90.9	1988	8	ATH131831	AJ131831 Arabidops
3	1804.4	90.9	2005	8	AY054480	AY054480 Arabidops
4	1748.4	88.1	1942	6	AX090345	AX090345 Sequence
5	1748.4	88.1	1942	6	AX090349	AX090349 Sequence
6	1747.6	88.0	1845	8	AF051849	AF051849 Arabidops
7	1747.6	88.0	1845	8	AF051849	AF051849 Arabidops
8	1034.4	52.1	1537	8	AF251794	AF251794 Brassica
9	954	48.1	1446	8	AF164434	AF164434 Brassica
10	652.8	32.9	2099	8	AF155224	AF155224 Arabidops
11	643.4	32.4	1964	8	AF129003	AF129003 Arabidops
12	592.6	29.9	92721	3	AC005917	AC005917 Arabidops
13	154.2	7.8	1497	3	AF221132	AF221132 Caenorhab
14	154.2	7.8	11966	3	CEH19N07	CEH19N07
15	145	7.3	1650	10	AF078752	AF078752 Mus muscu
16	145	7.3	1776	10	BC003717	BC003717 Mus muscu
17	142.6	7.2	2074	3	AY051835	AY051835 Arabidops
18	141.2	7.1	1976	3	AF059202	AF059202 Homo sapi
19	139.6	7.0	1493	9	BC006263	BC006263 Homo sapi
20	139.6	7.0	1895	6	AX090339	AX090339 Sequence
21	139.6	7.0	1998	9	BC015762	BC015762 Homo sapi
22	137.6	6.9	728	10	RNO345014	RNO345014 Rattus no
23	137.6	6.9	1751	10	AF296131	AF296131 Rattus no
24	137.6	6.9	1766	6	AX090340	AX090340 Sequence
25	136.4	6.9	1792	9	AF236018	AF236018 Cercopit
26	133.6	6.7	34948	3	CBRC33E23	CBRC33E23
27	129.8	6.5	253305	3	PFMAL3P7	PFMAL3P7
28	125.8	6.3	275	6	AX090352	AX090352 Sequence
29	122.6	6.2	234	6	AX090350	AX090350 Sequence
30	115	5.8	396	6	AX045795	AX045795 Sequence
31	109.4	5.5	254	6	AX090355	AX090355 Sequence
32	108.2	5.5	267	6	AX090351	AX090351 Sequence
33	103.6	5.2	262	6	AX090356	AX090356 Sequence
34	94.4	4.8	185932	2	AP003714	AP003714 Oryza sat
35	93	4.7	253	6	AX090354	AX090354 Sequence
36	86.4	4.4	40494	2	AC005122	AC005122 Arabidops
37	86.4	4.4	60144	2	AC019798	AC019798 Arabidops
38	86.4	4.4	91685	2	AC005121	AC005121 Arabidops
39	86.4	4.4	170362	3	AC009250	AC009250 Arabidops
40	86.4	4.4	258166	3	AE003652	AE003652 Arabidops
41	71.8	3.6	2072	8	SCU51790	SCU51790 Arabidops
42	71.8	3.6	2465	8	SCYNR019W	SCYNR019W Arabidops
43	68.6	3.5	2839	8	SCU5383	SCU5383 Arabidops
44	68.4	3.4	325	6	AX090357	AX090357 Sequence
45	68.2	3.4	1673	9	AF053336	AF053336 Chlorocob

ALIGNMENTS

RESULT 1	ATH238008	1904 bp	linear	PLN 18-JUN-1999
LOCUS	Arabidops	thaliana mRNA for diacylglycerol acyltransferase.		
DEFINITION	AJ238008			
ACCESSION	AJ238008.1	GI:5123717		
VERSION				
KEYWORDS	Dgat gene; diacylglycerol acyltransferase.			
SOURCE	thale cress.			
ORGANISM	Arabidops thaliana			
REFERENCE	1 (bases 1 to 1904)			
AUTHORS	Zou,J., Wei,Y., Jiao,C., Selvaraj,G. and Taylor,D.C.			
TITLE	The Arabidops thaliana TAG1 gene encodes for a diacylglycerol acyltransferase			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 1904)			
AUTHORS	Zou,J.			
TITLE	Direct Submission			
JOURNAL	Submitted (06-APR-1999) Zou J., Plant Biotechnology Institute, National Research Council of Canada, S7H 0W9, Saskatchewan, CANADA			
FEATURES	Location/Qualifiers			

source 1.1904
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
gene 139.1701
/gene="dgat"
CDS 139.1701
/gene="dgat"
/function="Lipid synthesis"
/product="diacylglycerol acyltransferase"
/protein_id="CA54537.1"
/db_xref="GI:5123718"
/translation="MALLDSAGVTTVTENGGEFVDLRLRRKSRSDSSNGLISGS
DNNSPSDVGAAPADVRDRIDSVDNDAGDTANLADNNGSGDNGGSGEGRGND
ATFTYRSPVPAHRRARRESPLSDAIFKQSHAGLFLNCVVLAVNSRLIENLMKGM
LIRDFWFSRSLRDMPLMCCIISLIPPLAFTVERKLYSEPVLEHITITM
TEVLVYVYTRCDASFLSGVITMLNCTIVLKIYSAHYSYDRLSANAADRANPEY
SYYSLSLAYFMAVPTLCOPSPRSCITKGVAROFALVIFGPMGITIEOYIN
PIVRSKRLPKGDLIAIERVLEKLSVPLVYVLCMFYCFELMINTILAEELCFEDREF
YKDMNNAKSVSDIYRMNMMPVHKMMVRIYFPLRSKIPKLAIIIFLIVSAVHELC
IAPCRLEKLMALFGLIMFQYPLVEITNLEDRFSGIVGMIFMFIPIGQPMCVLLY
YHDLNMRKGSMS"
BASE COUNT 458 a 399 c 431 g 616 t
ORIGIN
Query Match 91.3%; Score 1813; DB 8; Length 1904;
Best Local Similarity 95.9%; Pred. No. 0;
Matches 1904; Conservative 0; Mismatches 0; Indels 81; Gaps 1;

661 agtagactcatcatcgaaatacttatagaagatggttgatgcatcagaacggaattctcg 720
|||||
580 AGTAGACTCATCATCGAAATACTTATGAGATGTTGGTTGATCACAAGGATTTTGG 639
721 tttagtcaaatgcgtctgcagatctgacgcttccatgtgtgtatatacccttcgac 780
|||||
640 TTtAGTCAAGATCGCTCGAGATTTGGCCGCTTTTCATGTGTGTATATCCCTTTGCATC 699
781 ttccctgtgcctcttaccggttgagaatgtgacttgaagaataatcataagaacct 840
|||||
700 TTTCCTTTGGCTGGCTTTACGTTGAGAAATTTGTTACATTAAGAAATATATAGAACCT 759
841 gtgtcatcttcttcatatataatacacaagaggtttgtatccagttacgtc 900
760 GTTGTGATCTTTCTTCATATATATATATACCATGACAGAGGTTTGTATCCAGTTACGTC 819
901 accctaaggatgtagttctgtcttcttatacaggtgtcaacttggatcttccctacatc 960
|||||
820 ACCCTAAGGTGTGATTTCTGCTTTTATACAGGTGTCATTTGATGCTCTCCTACCTTCAT 879
961 gtgtgctaaagtgtgttcttatagtcatatagatagatataagatccctacgcaat 1020
|||||
880 GTGTGCTAAAGTTGGTTCTTATGCTCATATAGATATAGATATAGATATCCCTACGCAAT 939
1021 gcaagctataagcccaatcctcgaagctcctactaagcttgaagcttgagcatat 1080
|||||
940 GCAAGCTATAGGCCCAATCCGTAAGTCTCTACTACGTTAGCTTGAAGCTTGGCATAT 999
1081 tcaatgtctctcccatgtgttcttatagtcatatagatagatataagatccctacgcaat 1140
|||||
1000 TTCATGTGCTCTCCCATGATTTGTATATGCAAGCAAGTTATCCAGTTCTGCAATGATACG 1059
1141 aaggtgttggtgtgtgtcaatttgaaacgtgtcatatcaccggaattcaggaattc 1200
|||||
1060 AAGGTTGGGTGCTGCTGTCATTTGCAAAACGTGTCATATTCACCGGATTCATGGGATTT 1119
1201 ataatgaacaataataataatcctatgtcaggaactcaagaactccttgaagaagcat 1260
|||||
1120 ATATATGAACAAT 1179
1261 ctctcatatgtcatatgaagaaggtgtgaagcttcaagttcaaatatataatgtgtctc 1320
|||||
1180 CTTCATATATGCTATATGAAGAAGTGTGAAGCTTTAGTCCAAATTTATATGTGTGCTC 1239
1321 tgcattgtctactgtcttctcacccttgggttaaacatattggagagctctctgtctc 1380
|||||
1240 TGCATGTCTACTGCTTTCCACCTTTGGTTAAACATATTGGCAGACCTTCTGCTTC 1299
1381 gggggtcgtgaattctacaaagaattgttgaatgcacaaagtgttggagattactgaga 1440
|||||
1300 GGGGATCGTGAATTTCTACAAAGATTGGTGAATGCAAAAGTGTGGAGATTACTGAGAGA 1359
1441 atgtgaaatatacctgttcaataatggaatgttgaagaatataatcttccgtgtctgc 1500
|||||
1360 ATGTGGAATATGCCCTGTTCATATAATGATGTTGTCACATATATATATATATATATATAT 1419
1501 agcaagataccaagaacactgcgaattatcatatgtcttccctagctctgcagcttccat 1560
|||||
1420 ACCAAGATATCCAAAGACACTGCGCATTTATCATTTGCTTTCCATGCTGCGAGTCTTCAT 1479
1561 gagctatcatcgagcttccctgtctgtcttcaagctatggagcttcttcttggattatg 1620
|||||
1480 GAGCTATGATCGCAGTGTCTGTGCTCTTTCACAGCTATNGGCTTTTCTTGGGATTAATG 1539
1621 ttccagtgcttggcttcatcaacaataatcatcaagaagaagtttggctcaaggtg 1680
|||||
1540 TTTCAGTGTCTTTGGTCTTCATCACAACATATATACAGGAAGAGTTGGCTCAACGGTG 1599
1681 gggacaatgatctctcgtgtcatcttcttgaacttctggacaacgagatgtgtgtctct 1740
|||||
1600 GGGACATGATCTTGTGTTCTATCTTGTGATTTTGGACAAACGATGTGTGCTTCTT 1659

QY 1741 tattaccagacgtgtaacccgaaagatcgtatcgtgaacacactgttcaaaaat 1800
 |||||||
 Db 1660 TATTACCAGACCTGATGAAACGAAAGATGATGATGAAACAACTGTTCAAAAAT 1719
 |||||||
 QY 1801 gacttctcaacatcatgctcgtgtgagatcccgatgtgtgtgtgtgtgtgt 1860
 |||||||
 Db 1720 GACTTCTTCAACATCATGAGCTGCTGTGATCTCGTGTGATGTGTGTGTGTGT 1779
 |||||||
 QY 1861 gctaaacgacaaatagtgatataacatgaaagaaatagatgtgtgtgt 1920
 |||||||
 Db 1780 GCTAAACGCAAAATAGTGTATATACCATGAAAGAAAGAAATATAGTGTGTGT 1839
 |||||||
 QY 1921 tctgcaaaatttctgtgagacacgcaaacccgtgtgtgtgtgtgtgtgtgt 1980
 |||||||
 Db 1840 TCTGCAAAATTTTGTGTAGACACGCAACCCGTTGATTTGTGTGTGTGTGTGT 1899
 |||||||
 QY 1981 ggcgc 1985
 |||||||
 Db 1900 GCCGC 1904

RESULT 2
 LOCUS ATH131831 1988 bp mRNA linear PLN 10-JUN-1999
 DEFINITION Arabidopsis thaliana mRNA for diacylglycerol O-acyltransferase.
 ACCESSION AJ131831
 VERSION AJ131831.1 GI:5050912
 KEYWORDS diacylglycerol O-acyltransferase.
 SOURCE
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 1988)
 Hobbs, D.H., Lu, C. and Hills, M.J.
 Cloning of a cDNA encoding diacylglycerol acyltransferase from
 Arabidopsis thaliana and its functional expression
 FEBS Lett. 452 (3), 145-149 (1999)
 JOURNAL 9313150
 MEDLINE 2 (bases 1 to 1988)
 REFERENCE Hills, M.J.
 Direct Submision
 Submitted (22-DEC-1998) Hills M.J., Department of Brassica and
 Journal Oilseeds Research, John Innes Centre, Colney Lane, Norwich, NR4
 7UH, U.K.

FEATURES
 source Location/Qualifiers
 1..1988
 /organism="Arabidopsis thaliana"
 /sub_species="Columbia"
 /db_xref="taxon:3702"
 231..1793
 /EC_number="2.3.1.20"
 /product="diacylglycerol O-acyltransferase"
 /protein_id="CAB44774.1"
 /db_xref="GI:5050913"
 /translation="MAILDSAGVTVTENGGEFVDLRLRRKRSRSDSNGLLSS
 /DNSSPDVAGADADRIDSDVNDAGCTALADNNGGNGGNGGEGEGENAD
 ATFTTRPSVPAIRARARBSPLSDAIFKSHGLFALCVVILAVSRLLIENLMYCW
 LRTDFPWSRSLRDMPLRMCISLSTPLAFTYEKVLVLOKVISSEPVVITLHITM
 TEVLVYVYVTLRCDASFUSGVTLMLTCTVWLKLVIAHTSDISLANADKAPREV
 SYVSKSLAPVAPVATLCQPSYPSRACIRKQWAFKAVITGFMGFLIDBYIN
 PIVRSKSLAPGDLVATERYLVKLSVPLNVLWCMYCFEHLMLNTLAEELCEGREF
 YKDMNNAKSGVDYWRMMNMPVHKMVRHIFEPCLRSKLPKTLAIIIAELVASVFEELC
 IAVPCRLFKLMAFLGIMQVPLVFTITNYQERFGSTVGNMIFMFLFCIFGQPMCVLLY
 YHDLNKRKSSMS"
 ORIGIN
 BASE COUNT 478 a 423 c 436 g 651 t

Query Match 90.9%; Score 1805; DB 8; Length 1988;
 Best Local Similarity 95.9%; Pred. No. 0;
 Matches 1896; Conservative 0; Mismatches 0; Indels 81; Gaps 1;

QY 1 attcttagctctctccctcaatccgcctctccctccatctagatctgttccctt 60
 |||||||
 Db 93 ATTCTTACCTCTTCCCTCAATCCGCTCTTCCCTCTCATTAATGATCTGTTCCTTT 152
 |||||||
 QY 61 tcaattctctcgaatgctctcgaatctctcgaagcctctctccagcgtgtt 120
 |||||||
 Db 153 TCAATTTCTTCTGATGCTTCTGATTTCTGTGAGCGCTTTTCTCCGCGCTGTTT 212
 |||||||
 QY 121 cgtcaaacgcttccgaatgagatcttgaatctctcgtgcgttaactacaggtgag 180
 |||||||
 Db 213 CGTCAACCGCTTTTCGAATGCGATTTGTGATTTGCTGCGGCTTACATGAGTACGAG 272
 |||||||
 QY 181 aacggtgagcgagagatctgacatctgaagcttcgtcgaagaaatcgagatcgag 240
 |||||||
 Db 273 AACGTTGCGGAGAGATTCGTGATCTGATGATGATGATGATGATGATGATGATGAT 332
 |||||||
 QY 241 tctctcaacgagact 300
 |||||||
 Db 333 TCTTTACGAGACTTCT 392
 |||||||
 QY 301 gctcccgagagcttagagatgagatgagatgagatgagatgagatgagatgagat 360
 |||||||
 Db 393 GCTCCCGCCGCGCTTAGGATGCGATTTGATTCCTGTTTACGATGACGCTCAGGAGACA 452
 |||||||
 QY 361 gccaatctgacgagataataacggtggtgagataataacggtggtgagagagcgagc 420
 |||||||
 Db 453 GCCAATTTGGCCGAGATATTAACGGTGTGGGATTAATTAACGGTGTGGGATTAATTA 512
 |||||||
 QY 421 ggaagaaggaagaaacgcgagatgctgaatctgaatctgaatctgaatctgaatct 480
 |||||||
 Db 513 GGAGAAAGAGAGAGAAACGCCGATGCTTACGATGATGATGATGATGATGATGATGAT 572
 |||||||
 QY 481 cgaagggcgaag 540
 |||||||
 Db 573 CGAGAGGCGAG 618
 |||||||
 QY 541 ttattcaacctctgtagtagtcttattctgtaaacagtagatcatcatcgaaat 600
 |||||||
 Db 619 ----- 618
 |||||||
 QY 601 ctatgaagcgatcgagatgagatgagatgagatgagatgagatgagatgagatgag 660
 |||||||
 Db 619 -----AGACCCATGCGCGGATTTATTTCAACCTCTGTGTAGTCTTATTTGCGTAAAC 671
 |||||||
 QY 661 agtagactcatcatcgaataatctatgaagatgagatgagatgagatgagatgagat 720
 |||||||
 Db 672 AGTAGACTCATCATCGAATAATCTTATGAAATGATGATGATGATGATGATGATGATG 731
 |||||||
 QY 721 tttagtcaagatcgctgagagatgagatgagatgagatgagatgagatgagatgag 780
 |||||||
 Db 732 TTTAGCTCAAGATCGCTCGAGATTTGCGGCTTTTTCATGTGTGTATATCCCTTGCATC 791
 |||||||
 QY 781 ttcccttgagctgctctgaagatgagatgagatgagatgagatgagatgagatgag 840
 |||||||
 Db 792 TTTCTTTGGCTGCTTTACGTTGAGAAATTTGATTTGAGAAATTTGATTTGAGAAACT 851
 |||||||
 QY 841 gttgcatcttctcatatataatacaacatgaagaggtttgttccaggtttacgct 900
 |||||||
 Db 852 GTTGATCACTTCTTCAATATTAATTAACATGAGAGAGGTTTGTATCCAGTTTACGTC 911
 |||||||
 QY 901 acccgaagtgatgctgctcttcttcaaggtgctgctgctgctgctgctgctgctgct 960
 |||||||
 Db 912 ACCCTGAAGTGATTTGCTTTTATACAGGTGATCTTTGAGGCTCTCTACCTGCACTT 971
 |||||||
 QY 961 gttgctcaagttgtttcttctatgctcatctagatgacataaagttccctagccat 1020
 |||||||
 Db 972 GTGTGCTAAAGTGTGTTTCTTATGCTCATATGATGATTAAGATTCCTACGCAAT 1031
 |||||||
 QY 1021 gcaagctgaataagcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcga 1080
 |||||||
 Db 1032 GCAGCTGATTAAGCGCAATCTGAAGTCTCTCTACTACTAGCTTGAAGAGCTTGCGCATAT 1091
 |||||||

Db 144 CGTCAACGCTTTTCGAAATGCGGATTTTGATTCGCTGCGCTTACTACGTCAGGAG 203
 Qy 181 aacggtgacgagagatcctcgcagatctctgaagcttcgcgcagaaatcagatcagat 240
 Db 204 AACGGTGGCGAGAGTTCGTCGATCTTGATAGGCTTGTGACGGAATATGAGATCGGAT 263
 Qy 241 tctcttaacgact 300
 Db 264 TCTCTTAACGACTCTTCT 323
 Qy 301 gctcccgacgacttaagagatcagatctctctctctctctctctctctctctctctctct 360
 Db 324 GCTCCCGCGAGAGTTCGATCGGATTCGCTTGTAAACGATGACGCTCAGGAGACA 383
 Qy 361 gccaattgacgagagataataacgagtggtgcgataataacgagtggtgcgagagcgcgc 420
 Db 384 GCGAATTTGGCGGAGATTAATACGATGCGATTAATACGATGCGATTAATACGATGCGG 443
 Qy 421 ggaag 480
 Db 444 GAG 503
 Qy 481 cggag 540
 Db 504 CGGAGGCGGAG 549
 Qy 541 ttattcaact 600
 Db 550 ----- 549
 Qy 601 cctatgaag 660
 Db 550 -----AGAGCGATGCGGAGATTAATACGATGCGATTAATACGATGCGGATTAAT 602
 Qy 661 agtagactcatcagaaatct 720
 Db 603 AGTAGACTCATCAGAAATCTTATGAAGTATGATGATGATGATGATGATGATGATGATG 662
 Qy 721 ttaagtcaag 780
 Db 663 TTTAGTTCAGAGATCGCTGAGAGATGCGGCTTTTCAATGTTGATGATGATGATGATG 722
 Qy 781 tttcct 840
 Db 723 TTTCTCTTGGCTGCTTATGAGTTCAGAAATGATGATGATGATGATGATGATGATGAT 782
 Qy 841 gttgcatct 900
 Db 783 GTTGCATCTTCT 842
 Qy 901 accctaagtgagatct 960
 Db 843 ACCCTAAGTGATCT 902
 Qy 961 gttggtcgaagt 1020
 Db 903 GTTGGCTAAAGTTGTTCTTATGCTTATGCTTATGCTTATGCTTATGCTTATGCTTAT 962
 Qy 1021 gcaagcagataagcgaatct 1080
 Db 963 GCAAGCTATTAAGGCAATCTCTGAAGTCTCTACTAGCTTGAAGCTTGAAGCTTGAAG 1022
 Qy 1081 ttatgtgtcgtccacatggtlctctcaagcagatctctctctctctctctctctctctct 1140
 Db 1023 TTCTATGTTGCTCCCATATGTGTATACAGCAAGTTATCCAGCTTCTCATATACGG 1082
 Qy 1141 aaggt 1200
 Db 1083 AAGGTTGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1142
 Qy 1201 ataatagaacataataatct 1260
 Db 1143 ATTAATAGAACATATATTAATCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1202

Qy 1261 cttctctatgctatigaaagagtggtgaaagctctcagttccaaattatctgtgtgtc 1320
 Db 1203 CTCTATATGCTATTTGAAGAGTGTGAAGCTTTTCAAGTTTCAATTTATATGCTGCTC 1262
 Qy 1321 tgcattctactgtct 1380
 Db 1263 TGCATTTCTAGCTCTTCT 1322
 Qy 1381 gggagtcgtgaaatctcaaaagatgtgtgaaatgcaaaagtgtgtgaaatctcaaaag 1440
 Db 1323 GGGAGTCGTGAATTTCAAAAGATGTGTGAAGTGTGAAGTGTGTGAAGTGTGTGAAG 1382
 Qy 1441 atgtgaaatgct 1500
 Db 1383 ATGTGAATTTGCTCTGCTTCAATTAATGATGATGATGATGATGATGATGATGATGAT 1442
 Qy 1501 agcaagatacnaaagacactgcacatcaatctctctctctctctctctctctctctct 1560
 Db 1443 AGCAAGATTAACCAAGACACTGCGCATTAATTCATTTCTCTCTCTCTCTCTCTCT 1502
 Qy 1561 gactatgacatgcagatct 1620
 Db 1503 GAGCTATGCAATGCGATTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1562
 Qy 1621 ttcaagtgct 1680
 Db 1563 TTTAGGTGCTTGGCTTCT 1622
 Qy 1681 ggaacatgact 1740
 Db 1623 GGAACATGATCTTCT 1682
 Qy 1741 tattaccagacactgacgaacgaagagacagatgacgaacgaacgaacgaacgaacgaac 1800
 Db 1683 TATTACCAAGACCGATGAAACCGAAGAGATGATGATGATGATGATGATGATGATGAT 1742
 Qy 1801 gaacttctcaacatctatgactctctctctctctctctctctctctctctctctctct 1860
 Db 1743 GACTTCTTCAAAATCATATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1802
 Qy 1861 gctaaacgaacatagtgatataacatctgaagaagaagaagaagaagaagaagaaga 1920
 Db 1803 GCTAAACGCAAAATAGCTTATTAACATTAACGATGATGATGATGATGATGATGATG 1862
 Qy 1921 tctgcaaaatctgtgtgagagacagcaaacgctgtgtgtgtgtgtgtgtgtgtgt 1978
 Db 1863 TCTGCAAAATTTGTGTGAGAGACAGCGGACCGCTTGTGTGTGTGTGTGTGTGTGT 1920

RESULT 4
 AX090345 1942 bp DNA linear PAT 21-MAR-2001
 LOCUS AX090345
 DEFINITION Sequence 38 from Patent WO0116308.
 ACCESSION AX090345
 VERSION AX090345.1 GI:13444207
 KEYWORDS
 SOURCE
 ORGANISM
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE
 1 (bases 1 to 1942)
 AUTHORS Lassner, M. and van Eenennaam, A.
 TITLE Plant sterol acyltransferases
 JOURNAL Patent: WO 0116308-A 38 08-MAR-2001;
 MONSANTO COMPANY (US)
 FEATURES
 source location/Qualifiers
 1..1942
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 BASE COUNT 461 a 421 c 425 g 635 t
 ORIGIN

SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
BASE COUNT
ORIGIN

thale cress.
Arabidopsis thaliana
Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Eukaryota; Viridiplantae; Magnoliophyta; eudicotyledons; core eudicots;
Spermatophyta; Eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1942)
Kasner, M. and van Eenennaam, A.
Plant sterol acyltransferases
Patent: WO 01/6308-A 42 08-MAR-2001;
MONSANTO COMPANY (US)
location/Qualifiers
1. 1942
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"

Query Match 88.1%; Score 1748.4; DB 6; Length 1942;
Best Local Similarity 95.7%; Pred. No. 0;
Matches 1840; Conservative 0; Mismatches 1; Indels 81; Gaps 1;

QY 1 attcttagctctcctcctcaatccgcctcttccctccatagatctgttccctt 60
DB 99 ATTCTTAGCTCTCTCTCAATCCGCTCTTCCCTCCATAGATTCTGTTCCCTCT 158
QY 61 tcaattctctgcatgctctcgaatctctctcgaagcctcttctccgaagctgtt 120
DB 159 TCAATTCTCTTGCAATTCGATTCCTGCTGACGCTCTTTCGCCGACGCTGTT 218
QY 121 cgtcaaacgctcttcgaatagcgatcttgatctctgctgctgctgctgctgctg 180
DB 219 CGTCAAACTCTTTCGAATTCGATTCCTGCTGACGCTCTTTCGCCGACGCTG 278
QY 181 aacgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 240
DB 279 AACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 338
QY 241 tcttctaagcagctctctctctctgctgctgctgctgctgctgctgctgctg 300
DB 339 TCTTCTAACGCACTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 398
QY 301 gctccgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 360
DB 399 GCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 458
QY 361 gccaatctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 420
DB 459 GCCAATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 518
QY 421 ggaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 480
DB 519 GGAAGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 578
QY 481 cgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 540
DB 579 CGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 624
QY 541 ttattcaacccctgctgctgctgctgctgctgctgctgctgctgctgctgctg 600
DB 625 ----- 624
QY 601 ctatgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 660
DB 625 -----AGAGCATGCGCGATTAATCAACCTCTGTAGTAGTTCTTATGCTGTAAC 677
QY 661 agtaaacatcatcgaataatctatgaagatgctgctgctgctgctgctgctgctg 720
DB 678 ACTAAGCTCATCATGAATAATCTTATGAAGTAGTGTGCTGCTGCTGCTGCTGCTG 737
QY 721 tttagtcaagaatcgtgctgctgctgctgctgctgctgctgctgctgctgctgctg 780
DB 738 TTTAGTCAAGATCGCTGCGATGCGCTTTTCATGCTGTATATATCCCTTTCATC 797

QY 781 tttccttgctgctcttcaagctgctgctgctgctgctgctgctgctgctgctgctg 840
DB 798 TTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 857
QY 841 gtttcaatcttcttcaatcttcaatcttcaatcttcaatcttcaatcttcaatctt 900
DB 858 GTTGTCACTTTCTTCAATATTAATCAACATGAGAGGTTTGTATCACTTACGTC 917
QY 901 accctaagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 960
DB 918 ACCCTAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 977
QY 961 gttgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1020
DB 978 GTTGTCTAAGGTTGCTTCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1037
QY 1021 gcaagctgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1080
DB 1038 GCAGCTGATTAAGCCCAATCTGGAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1097
QY 1081 tctatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1140
DB 1098 TTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1157
QY 1141 aagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1200
DB 1158 AAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1217
QY 1201 ataataagaacatataataatcttctgctgctgctgctgctgctgctgctgctgctg 1260
DB 1218 ATAAATGAACATATTAATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1277
QY 1261 ctctataagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1320
DB 1278 CTCTATATGCTATTAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1337
QY 1321 tgcagcttctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1380
DB 1338 TGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1397
QY 1381 gggagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1440
DB 1398 GGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1457
QY 1441 atgtggaatctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1500
DB 1458 ATGTGGAATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1517
QY 1501 agcaagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1560
DB 1518 AGCAAGTATCAAAAGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1577
QY 1561 gaagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1620
DB 1578 GAGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1637
QY 1621 tttaagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1680
DB 1638 TTTCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1697
QY 1681 ggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1740
DB 1698 GGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1757
QY 1741 tattaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1800
DB 1758 TATTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1817
QY 1801 gaacttcttcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1860
DB 1818 GACTTCTTCAAAACATCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1877

QY 1861 gctaaacagcaaatgtgttataaccattgaagaagaagaattaggtgttga 1920
 Db 1878 GCTAAACCAACAAATAGTGTATAACCATTTGAAGAAAGAACAACTTAGAGTTGTGA 1937
 QY 1921 tc 1922
 Db 1938 TC 1939

RESULT 6

AF051849

LOCUS

1845 bp mRNA linear PLN 24-JAN-2000

DEFINITION

Arabidopsis thaliana diacylglycerol acyltransferase (DAGAT) mRNA,

ACCESSION

AF051849 AF061563

VERSION

AF051849.1 GI:6625552

KEYWORDS

thale cress.

SOURCE

Arabidopsis thaliana

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

REFERENCE

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

AUTHORS

Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

TITLE

Bouvier Nave, P., Benveniste, P., Oelkers, P., Sturley, S.L. and

JOURNAL

Schaller, H.

MEDLINE

Expression in yeast and tobacco of plant cDNAs encoding acyl

REFERENCE

Eur. J. Biochem. 267 (1), 85-96 (2000)

AUTHORS

2 (bases 1 to 1845)

TITLE

Benveniste, P.

JOURNAL

Direct Submission

COMMENT

Submitted (28-FEB-1998) Institut de Biologie Moleculaire des

FEATURES

Plantes, Centre National de la Recherche Scientifique, 28 rue

source

Goethe, Strasbourg 67083, France
 On Jan 24, 2000 this sequence version replaced gi:6650202.

Location/Qualifiers

1..1845

organism

"Arabidopsis thaliana"

cultivar

"Columbia"

db_xref

"taxon:3702"

tissue

"type="hypocotyl"

clone

"lib="ABRC library, catalog number CD4-15"

dev

stage="3 day-old seedling"

note

"screened with EST, Genbank Accession Number

AA042298"

1..1845

gene

/gene="DAGAT"

CDS

88..1650

function

"catalyzes the acylation of diacylglycerol by

oleoyl-CoA

to give triacylglycerol"

product

"diacylglycerol acyltransferase homolog"

protein

id="AA19262.1"

db_xref

"GI:6625553"

translation

"MatisAGVTVTENGGEFVDLRLRRKSSDSSNGILLSGS

DMSPPDVGADPADVDRIDSVVNDAGTANLAGDNGGDDNGGREGGREGNAD

LIPDPWFSSRLDMPLEMCISLIFPLAFTVRLVLOKTSPPVYFELIITM

TEVLPYVTLKCDASFLSGVTLMLTCTVMTKLVSVAHSTYDTSRLAANAARANEV

SYVYSLSLAFVAVAPLQPSYPSRACIRKQWAFROKAKIVTFEPGFTTEOYTIN

PIVRNSKHLPLKGLLAIERVLKLSVPTNLYVMICMVCYCFPHMLNLIALDLGDEF

YKDMNAKSYGDYRWMMNVHKKMVRHIVPRLSKIRPKTLAIILVSAVPELCL

IAYPCFLPKIMAFILGIMPOVPLVFTITNLDERGSTVGMFLFWFICLFGQPMCVILY

YHDMNRKMSMS

BASE COUNT

453 a 378 c 421 g 593 t

ORIGIN

Query Match 88.0%; Score 1747.6; DB 8; Length 1845;
 Best Local Similarity 95.6%; Pred. No. 0;

Matches 1841; Conservative 0; Mismatches 4; Indels 81; Gaps 1;

QY 52 tttctcttcaattcttcttgcagtcgtcttcgattctctctgacgcctcttctccg 111
 Db 1 tttctcttcaattcttcttgcagtcgtcttcgattctctctgacgcctcttctccg 60
 QY 112 aagctgttctcgaacgcttcttcgaattgagatttggattctctgagcttactag 171
 Db 61 ACCGCTTTTCGTCGTAACAGCTTTTCGAAATGCCATTTTGGATTTCTGCGCTTACTACG 120
 QY 172 gtaacggaagaacgctgagcgaagtcgttcgattctgataagcttcgacgaagacg 231
 Db 121 GTGACGGAAGACGCTGCGGAGAGTTCGATCTTGATAGGCTTGCGAGGAATATG 180
 QY 232 agatcgattctctcgaacgactcttctctggttcgataaattctcttcgagat 291
 Db 181 AGATCGGATTCCTTAACGAGACTTCTCTGTTCCGATTAATTAATTCCTTCGGAT 240
 QY 292 gatgttgagatcttcgacgagcgaagtcgattgattcttcgttgtaacgagatcgt 351
 Db 241 GATGTGGAGCTCCCGCCGACGTTAGGATTCGATTCGATTCGATTCGATTCGATTC 300
 QY 352 cagggagacagcgaattctgacgagataaataacgctggtgagataaataacggtgga 411
 Db 301 CAGGGAACAGCCAAATTTGGCCGAGATTAACGGTGGGAGATTAACGGTGGGAG 360
 QY 412 aagagcgcgaggaaggaaggaaggaacgcgaagtcgattacgttcgacgagtcgt 471
 Db 361 AGAGCGGCGGAGGAAGGAAGGAAGGAAGGACCGCATCTTACGTTACGTTACGTTAC 420
 QY 472 ccagcttcaggaagcagagagagagagtcacgttccttcgacgacgaattcttaacagagc 531
 Db 421 CCAGCTCATCGAGCGGAGAGAGAGAGTCACTTACCTCCGCAATCTTCAAC----- 475
 QY 532 catgacgattatcaaacctctgtgtagtcttcttctgctgaacagtagatcctc 591
 Db 476 ----- 475
 QY 592 atcgaataatcttgaagagcagtcgagattatcaactctgctgtagtcttcttatt 651
 Db 476 -----AGAGCATGCGCGGATTTATCAACCTCTGTGTAGTATCTTATT 519
 QY 652 gcttcaacagtagtactcaatcaatcgaataatcttgaagtagtctgtagtcagagag 711
 Db 520 GCTGTAAACAGTACACATCATCATGAAATCTATGAAGTAGTGTGATGATCAAGC 579
 QY 712 gattcgtgtttagtcaacagcgtcgcgagatctgacgcttctcaatggtgtagatcc 771
 Db 580 GATTCTGTGTTAGTTCAACATGCTGCGAGATTCGCTTTCATGTGTGTATATCC 639
 QY 772 ctttcgactcttcttgcgtcgtcttcaagctgtaagaattgtagtctagaatacata 831
 Db 640 CTTTCGATCTTCTTCTTGGCTGCTTACGGTTCGAAATTTGGTACTTCAGAAATACATA 699
 QY 832 tcgaacactgttcaatcttcttcaatattatcaatcaatgaacgaaggttctgatacca 891
 Db 700 TCAGAACCTGTTGTCATCTTCTTCAATATTATATACCATACAGAGGTTTGTATATCA 759
 QY 892 gtttaagctaaccttaagctgtagtcttcttcttcaatgaagctgtagtctgctc 951
 Db 760 GTTTAGCTACCTTAAGGTGATCTGCTTTTATATGAGTGATCTTATGATGCTCTC 819
 QY 952 acttgatgctgtgctgaagctggttcttcttcaatgaatgtagtctgtagtctgtagc 1011
 Db 820 ACTTGATGCTGTGCTGAAGGTGTTCTTATATGCTATAGTATGATATGATATGATC 879
 QY 1012 ctgacgaatgacgtgataagcgaatcctgaagctctctactactgtagtctgtagc 1071
 Db 880 CTAGCCAATGACGTGATTAAGGCCAATCTGAAAGTCTCTTCTAGTATGCTTAAAGC 939
 QY 1072 ttggatattcagtgctgctcccaatctgcttcaatgaacgaattcaacgcttctgca 1131
 Db 940 TTGGCATATTTCATAGTGTGCTCCACATTTGTTATACAGCAAGTATACAGTTCATGCA 999

QY 1132 tgatagaagaggttggtggtcgtcgaatttcgaacacgtgcatatccaccgagctc 1191
 |||||||
 Db 1000 tgrataggaaggttggtggtcgtcgaatttcgaacacgtgcatatccaccgagctc 1059
 QY 1192 atgggatttataagaacaatatataatccattgtcagagacccaagaccccttg 1251
 |||||||
 Db 1060 atgggatttataagaacaatatataatccattgtcagagacccaagaccccttg 1119
 QY 1252 aaagggactctctataatgcatatgaagagtgtaagcttcagttcccaattatlat 1311
 |||||||
 Db 1120 aaagggactctctataatgcatatgaagagtgtaagcttcagttcccaattatlat 1179
 QY 1312 gttggtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1371
 |||||||
 Db 1180 gttggtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1239
 QY 1372 ctctggtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1431
 |||||||
 Db 1240 ctctggtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1299
 QY 1432 tactggaagatgtaagatgtaagatgtaagatgtaagatgtaagatgtaagatgtaag 1491
 |||||||
 Db 1300 tactggaagatgtaagatgtaagatgtaagatgtaagatgtaagatgtaagatgtaag 1359
 QY 1492 tggcttgcgcagcaagaataccaaagaacacgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1551
 |||||||
 Db 1360 tggcttgcgcagcaagaataccaaagaacacgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1419
 QY 1552 gttcttcaatgagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1611
 |||||||
 Db 1420 gttcttcaatgagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1479
 QY 1612 gggatattgtttcaggtggtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1671
 |||||||
 Db 1480 gggatattgtttcaggtggtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1539
 QY 1672 tcaagagtggtggaagacatgacatgacatgacatgacatgacatgacatgacatgacatgacat 1731
 |||||||
 Db 1540 tcaagagtggtggaagacatgacatgacatgacatgacatgacatgacatgacatgacatgacat 1599
 QY 1732 gttgtcttcttataccacgacacgtgacacgacacgacacgacacgacacgacacgacacgt 1791
 |||||||
 Db 1600 gttgtcttcttataccacgacacgtgacacgacacgacacgacacgacacgacacgacacgt 1659
 QY 1792 tcaaaaaatgacatttcttcaaacatctatgacacgtgacacgtgacacgtgacacgtgacacgt 1851
 |||||||
 Db 1660 tcaaaaaatgacatttcttcaaacatctatgacacgtgacacgtgacacgtgacacgtgacacgt 1719
 QY 1852 gttctgagatgtaagacgaacatgtaagacgaacatgtaagacgaacatgtaagacgaacatgtaag 1911
 |||||||
 Db 1720 gttctgagatgtaagacgaacatgtaagacgaacatgtaagacgaacatgtaagacgaacatgtaag 1779
 QY 1912 gttgtgtatctgcaaaaaatttggtagagacgaacacccgttggtagatcttggtagatg 1971
 |||||||
 Db 1780 gttgtgtatctgcaaaaaatttggtagagacgaacacccgttggtagatcttggtagatg 1839
 QY 1972 ttttaa 1977
 |||||||
 Db 1840 ttttaa 1845
 |||||||
 RESULT 7
 AF251794 1537 bp mRNA linear PLN 16-APR-2000
 LOCUS AF251794 Brassica napus putative diacylglycerol acyltransferase mRNA,
 DEFINITION complete cds.
 ACCESSION AF251794
 VERSION AF251794.1 GI:7576940
 KEYWORDS
 SOURCE
 ORGANISM
 Brassica napus
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
 REFERENCE
 AUTHORS Brown, A.P., Schlerer, T.P. and Slabas, A.R.
 TITLE Characterization of a putative diacylglycerol acyltransferase cDNA
 from Brassica napus embryo
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1537)
 AUTHORS Brown, A.P., Schlerer, T.P. and Slabas, A.R.
 TITLE Direct Submission
 JOURNAL Submitted (04-APR-2000) Biological Sciences, University of Durham,
 South Road, Durham DH1 3LE, UK
 FEATURES
 source
 1..1537
 /organism="Brassica napus"
 /cultivar="Commanche"
 /db_xref="taxon:3708"
 /dev_stage="embryo"
 22..1527
 /EC_number="2.3.1.20"
 /product="putative diacylglycerol acyltransferase"
 /protein_id="AF251794.1"
 /db_xref="GI:7576941"
 /translation="MEILDSGVTPTEMGADLDLRLHRRPRSDSSNGLLPDSVTVS
 DADVDKDSAVEDTQGRKANLAGEEISEGEGANVDVRYTRPSVPAHRRVRES
 LSSDAIFKOSHAGLFNLCVYVAVNSRLILENLMKYGMLIRTFDWFSSLRDPLP
 MCCSLSTFPLAFTVEKLVLOKCSSEPVVILIIITMVELVPPVYTLKDSALFS
 GYTLMLILCTVWKLKYSVAHRTDRIYRLANSDDKRNPEVSYVSLKAPMLPTLC
 YQSYPRSPCINKGVNARQFAKLIITFGPMGFIIEQYINPIYVRNSKHPKGDILXGE
 RVKLKSVPLVYVLCMYCFEFLMNLNLAELLCFEDREYIDDMNKAQSVGYRRMNA
 PVKLMVRRVYPCRLRNIPKVPAILIIFLVSAYVHELCLIAVPCFLKLMFLGIMFO
 VPLVFITNLOERFESMVGNMIFMTFFIFGPMKVLVYHDLNMRKMS"
 BASP COUNT 371 a 322 c 355 g 489 t
 ORIGIN
 Query Match 53.2%; Score 1055.4; DB 8; Length 1537;
 Best Local Similarity 81.9%; Pred. No. 1.2e-259;
 Matches 1350; Conservative 0; Mismatches 161; Indels 138; Gaps 6;
 QY 132 ttccgaatgagcatttgatcttctctctgctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 191
 |||||||
 Db 15 ttttccaatgagcatttgatcttctctctgctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 74
 QY 192 agaagtcgacatcttgatagcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 251
 |||||||
 Db 75 -----GATCTGATACCGCTTCGTACCGGAACCGAGATCGGATCTTCCAATGG 125
 QY 252 actctctctctctggttcgataataatctctctctctctctctctctctctctctctctctctcgcga 311
 |||||||
 Db 126 ACTCTCTC---CTGATTCCTGTAACCTGTTCCGAT-----GCTGA 161
 QY 312 cgttaagagatcgatctgattcccgltgttaacgagcgtcgaaggaacccaatttggc 371
 |||||||
 Db 162 cgttaagagatcgatctgattcccgltgttaacgagcgtcgaaggaacccaatttggc 218
 QY 372 cgaagataataacgltgttgcaataacgltgttgcaagagcggtcgaaggaaggaag 431
 |||||||
 Db 219 CGGAG-----AAACGAATATAGGGAATCCCGTGAGAGAGCGG 257
 QY 432 aggaagcgcgatgctacgtttagcgttcgacgttcggttcgacgttcgacgttcgacgttcgacgttcgacgttc 491
 |||||||
 Db 258 GGGAAACGTGATGTAAAGTACAGCTATCGGCTCGCTTCACCTCAGTACGAGAGGATGCG 317
 QY 492 agagagtcacattgctcgcagcgaacatctcaacagagcaatgctcggtattatcaact 551
 |||||||
 Db 318 GGAAGTCCACTCAGCTCTGACGCGCATCTTCAAC----- 352
 QY 552 ctgtgtagtagttcttattctgttaaacagtagactcatcatcgaaaaattatgaagag 611
 |||||||
 Db 353 -----AGAG 356

```

QY 612 ccatgcccgaattatcaacctctgtgagtagtcttatctgtcgtgtaaacagtagactcat 671
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 357 CCATGCGGACATTCACACCTGTGTAGTAGTTCTTGTCGTCTTAACAGTAGACATCAT 416
QY 672 catcgaataactcttaagaatgtagtctgtatcagaagaagattctctgtttagtcaag 731
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 417 CATCGAATAACTCATGAGTAGACGTTGGTGTGATCAGAACTGATTTCTGTTAGTTCAAC 476
QY 732 atcgcgcgagattggccgcttctcatgtgtgtatatacccttctgaacttctcttggc 791
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 477 GTCCTGCGAGGATTTGGCCCTTTTTCATGTGTGTCTCTCCCTTTCATGTTCTTCCCTTGGC 536
QY 792 tgccttcaaggttgagaatgtgtacttcgagaataataatcagaacctgtgtacatct 851
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 537 TGCCTTACCGCTTCGAGAAATAGTACTTGAATATGATATGTAACCTGTTCATCAT 596
QY 852 tctcatatataatcacatagacaaggtttgtatccagtttaacgtcaacctaaagtg 911
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 597 TCTTCATATTTATATACCATGACGAGGCTTGTATCCAGTGTATGCTACCTGAAGCTG 656
QY 912 tgaattcgtctttaaatacaggtgtcacttgatgcctcacttgatgtgtgtgctaa 971
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 657 TCATTCGCGCTTCTTATCAGGTGTACGTTGATGCTCTCCTACCTGATGTGTGCGCTGAA 716
QY 972 gttgtgttcttctgtcactactagctatgaataagaatccctagccaatgcaagtata 1031
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 717 GTTGGTTTCTTACGCTACTACTATACATTAACAAACCCCTAGCTAATTCATCTATATA 776
QY 1032 ggcacatcctgaagctcctactacgtactgtagtgaagagcttggaataatctatgc 1091
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 777 GGCACATCTGAAAGTCTCTACTATGTAGTTGAAGACCTTGCGCTATTTCATGCTTGC 836
QY 1092 tcccatgtgttatacagccaagtataccacgtctcgtcatatatacgaagagttgggt 1151
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 837 TCCCATTTGTATACACCGACGATATCCACCTTCTCCATGTATCCGGAAGGTTGCGT 896
QY 1152 ggcctgcgaatttgcaaaactgtcatatctacccgagatcaggaattataataga 1211
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 897 GGCCTGTCATTTGCAAGCTGATCATATTCCTGATTCATGGGATTTATATAGAGCA 956
QY 1212 atataataactatctgtcagaactcaaaagcactcttgaaaggcgactctctatagc 1271
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 957 ATATATAATCTATATTTGTAGGAACTCAAAACATCTTGAAGAGGATCTCTTATACG 1016
QY 1272 tatgaagaagtggtgaagcttgcagttccaaattatatgtgtgtctgtcgaagtctta 1331
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1017 TGTTAAGAGGTTGAGAGCTTTCAGTTCCAAATTTATACGTGCTGTCATGTTCTTA 1076
QY 1332 ctgcttcttcaaccttggtaaacatactggcagaagctctctgtcgcggaatcgtga 1391
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1077 CTGCTTCTCCACCTTTGGTTAAACATATTGACAGAGCTCCTGCTTGCGGATCGTGA 1136
QY 1392 attcaaaaagattgtggaatgcaaaaagtgtggaatactactggaagaatgtgaatac 1451
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1137 ATTCTACAAAGATTTGGTGAATGCAAAAAGCGTGGAGATTATTTGGAGAAATGTGGAATAT 1196
QY 1452 gctcgttcaataaagatggtgcacataatactccggtgtcgtgcgacaagaatacc 1511
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1197 GCTGTGTATTAATGAGATGTTGCGATGATTAATCTTCGATGCTTCCGACAAAATATACC 1256
QY 1512 aaagaacatcgcaattatcaattgcttctagctcgtcagctccttcaatgaactatcat 1571
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1257 GAAAGTACCCCGCTATTATCCCTTGTAGTCTCTGCTCACTTTCATGATGATATGATCAT 1316
QY 1572 cgaagcttctgtcgtccttcaagctataggtcttctgtggaattatgttcaggtgcc 1631
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1317 CGCAGTTCCTTGTGTCTCTTCAACTATGCGCTTCTTCTTGGGATTTATGTTTACGCTGCC 1376
QY 1632 ttgtgtctcaatcaacaactatactacaggaagaagtttgctccaaggtgtgggaacatgat 1691
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1377 TTTGGTATTATACAAACTACCTACAAAGAAAGTTTGGCTCCATGTGTGGGAAACATGAT 1436
QY 1692 ctctcgttcaatctctcgaacttctcgaaacagatgtgtgtcttcttatttaaccaaga 1751

```

```

Db 1437 ATTCGTGTTTACCTTTCGATTTTCGACAAACCGATGTGTGCTTCTTATATACGA 1496
QY 1752 cctgatgaaccgaagaagatcatgcat 1780
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1497 CTTCATGACACCGCAAGAAGATGATCAT 1525

RESULT 8
AF164434 1512 bp mRNA linear PLN 30-NOV-1999
LOCUS Brassica napus putative diacylglycerol acyltransferase (DGAT1)
DEFINITION mRNA, complete cds.
ACCESSION AF164434
VERSION AF164434.1 GI:5579407
KEYWORDS rape.
SOURCE Brassica napus
ORGANISM Brassicaceae: Brassica napus
REFERENCE 1 (bases 1 to 1512)
AUTHORS Nykiforuk, C.L., Larchoe, A. and Weselake, R.J.
TITLE Isolation and Characterization of a cDNA Encoding a Second Putative Diacylglycerol Acyltransferase from a Microspore-derived Cell Suspension Culture of Brassica napus L. cv Jet Neuf (Accession No. AF164434). (PGR99-158)
JOURNAL Plant Physiol. 121 (3), 1053 (1999)
REFERENCE 2 (bases 1 to 1512)
AUTHORS Nykiforuk, C.L., Larchoe, A. and Weselake, R.J.
TITLE Direct Submission
JOURNAL Submitted (30-JUN-1999) Chemistry and Biochemistry, University of Lethbridge, 4401 University Drive, Lethbridge, Alberta T1K 3M4, Canada

FEATURES
Source location/Qualifiers
1..1512 /organism="Brassica napus"
/db_xref="taxon:3708"
1..1512 /note="microspore-derived cell suspension culture"
1..1512 /gene="DGAT1"
1..1512 /gene="DGAT1"
/EC_number="2.3.1.20"
/codon_start=1
/product="putative diacylglycerol acyltransferase"
/db_xref="GI:5579408"
/db_xref="AF164434.1"
/translation="MALLDSGVAVPPTENGVAJDLRLHRRKSSDSSNGILSDTSPS
DDVGAIAERDVRDVAEEAEAGGTANLAVAGDAEERESAGGVRTYRPSVPAHRTRE
SPLSDAIFKOSHAGFLNLCVYVAVANSLRIENLKKYGLIRDFEFSSTLRDMP
LFMCLSISVPLAFTVEKAVLQKFTSEPAIILHYITPTEVLYVYVYTLRCDSPF
LSGVTLMLNCIIVWLKLYSVASVYDIIRTLANSADKDPRETSVYSLSKSLATPVNATP
LCVQSPVPRSPKIRKGVAVARLALVLTFTGLMGITIOYINPIYVNSRHPKGLDLYA
IERVLSVENVLYVDMTCFELHMLNLLAEILCFEDRETYKDMNNAKSVGDVWRW
NMPEVHKMVRHVPICRIKIPKPAIILFVLVASVHELCIAVPCILFNLMEGMIT
FVQVPLVITITNLEDFRGSVMWNAACGFCFQPMGLLYHDLNMRKMSMS"
BASE COUNT 344 a 343 c 363 g 462 t
ORIGIN

Query Match 52.1% Score 1034.4 DB 8 Length 1512:
Best Local Similarity 80.4% Pred. No. 2.9e-254:
Matches 1321; Conservative 0; Mismatches 191; Indels 132; Gaps 4;
QY 139 atggcgaatttggaattcgtcgtcgttcaacggtgacggaacggtggcggaagttc 198
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 ATGGCGATTTTGGATTTCGAGGCGTGTACCGCGGAGGAGAAACGGCG-----TC 54
QY 199 gtccatctgaagatctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 258
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 55 GCGGATCTCGACAGCGTCCACCGTCGTAATACGAGTTTCGATTCGCAACGAGCTCTC 114

```


/db_xref="GI:5225382"
 /translation="MCCSLSTFPLAETVEKIVLQKCISEPVVIFLHVITMTEVLY
 PYVYLRDCAFLSGDTMLLCTIWLIVLIVSAHNVARQKIFLGLMSEFIIIOYINPIVSN
 LKSLAYFMILAPLYQSPSPCKRMVVARQKIFLGLMSEFIIIOYINPIVSN
 SKHPLKGLDLVLAIEVRLKSLVSNLYWLCMEYCEFHMLNLIALLCEFGDEEYKDMW
 NAKSYGDIWRKMMNMFHKNMVRHNVFPLKRIIPVPAIIITAFIVSAFHELCIAVPC
 RLNFNFMHGLMFVPLVFTNFLOERBSVMVMIFGASCIKFGPWCGLIYHDLML
 NRKMSM"
 BASE COUNT 398 a 291 c 278 g 479 t
 ORIGIN

Query Match 48.1% Score 954; DB 8; Length 1446;
 Best Local Similarity 91.8%; Pred. No. 1.2e-233;
 Matches 1019; Conservative 0; Mismatches 90; Indels 1; Gaps 1;

QY 675 cgaatacttataagatgtgtgtgtatcatcagaacgattcttggttaattcaagtc 734
 DB 1 CGAAATATCATGAGTACGAGTGTGTGATCAGAACTATTTCTGTTAGTCAACGTC 60
 QY 735 gctgagattggcgcttttcattgtgtatatacccttcgatcttcccttgctgc 794
 DB 61 GCTGCGAATTT-GCCGCTTTTCATGTTGTCCTCCCTTCAATCTTCTTGCGCTGC 119
 QY 795 cttaacgttgagaatgtgtacttcaagaatatacagaacctgtgtcatcttctc 854
 DB 120 CTTAACGCTGAGAAATGATCTTCAAGAAATGATATCTGAACTGTTGATCTTCT 179
 QY 855 tcatatataataccatgacagaagtttgcataccagttacagccctaaagttga 914
 DB 180 TCAATGATTAATWACATGACCGAGGCTGTGTATCCAGTATATGCTCAAGAGTGA 239
 QY 915 ttcgcttttatacaggtgtcacttgatgctcctcacttgcattgtgtggttaagt 974
 DB 240 TTCTGCTTTTATACAGTGTACAGCTTGAGCTGCTTCACTGATGTTGCTGGAAGTT 299
 QY 975 gtttcttatagtcactactgatacagaagaatccctagccaatgcaatgataagc 1034
 DB 300 GGTTCCTTACGCTCATCTACTAATGACATGAAGAACCCCTAGTAATCATCTATAAGC 359
 QY 1035 caatccgaagctcccaactgaactgaactgaagagcttgacattatcaatgacatc 1094
 DB 360 CAATCCGAGAGTCTCCACTATGTAGCTTGAGAGCTTGCTTATTTTCAATGCTCTCC 419
 QY 1095 caattgtgtatcagcaagatataccagttcgtatgatalacaggaaggtgtgtgc 1154
 DB 420 CACATGTGTATACGCAAGCTATCCAGCTCTCCATGATGATCCGGAAGGTTGGCTGC 479
 QY 1155 tgcataattgcaaaactgtcatatcccgatcatcggaattataatgaacaata 1214
 DB 480 TCGTCAATTTGCAAACTGCTCATTTTCACTGACATGAGGATTTAATAGACAAATA 539
 QY 1215 tataatcctatgtcagaactcaagcatccttgaagagcgatcttcatatgctat 1274
 DB 540 TATAAATCTATTTAGAGAACTCAAAAGCATCTCTGAAAGGAGCCTTCTATATGCTAT 599
 QY 1275 tgaagaaggttggaacttgatcagttcccaattataatgtgtgctgtcatgttact 1334
 DB 600 TGAAGAAGTGTGAAAGCTTCAAGTTCAATATATGTGTGCTGTGATGCTCTACAG 659
 QY 1335 ctcttccaccttggttaacataatgacagagcttctgctcgggagtcggaatt 1394
 DB 660 CTTCCTCCACCTTTGTTAAACATTTGGCAGAGCTCTCTGCTTGGGGACCGTGAAT 719
 QY 1395 ctacaaagattggtggaatgcaaaagtgtggaatatactactggaagatggaatgcc 1454
 DB 720 CTACAAAGATTGCTGGAATGCAAAAAGCTGAGATATTGGAAGATGTGAATATGCC 779
 QY 1455 tgtataataagatggtgtgacatatatactcccggtctgtgcgacgaagataccaa 1514
 DB 780 TGTTCACAAATGAGATGTTCAATGATATCTTCTGCTGCTGCAACAGATACCAAA 839
 QY 1515 gaacactgcacatcatcattgtcttctagctctgcagccttcatgacatgacatgc 1574

DB 840 AGTACCCGCCATWATCATTTGCTTCTTATAGTCTGACAGTCTTTCATAGTTATGATCGC 899
 QY 1575 agttcctgtgtctcttcaagcatgagcttctctgggattatgttcaagtgcttc 1634
 DB 900 AGTTCCTTGCGCTGCTTCATCTATGAGCTTTCATGGAATATATGTTGAGTCCCTTT 959
 QY 1635 gttctcatcacaaactatacagaagaagtttgctcaagcgttgaggagacatact 1694
 DB 960 GGTCTTATACAAACATTTTACAGAAAGTTTGGCTCATGCTGGAACATGATCTT 1019
 QY 1695 ctggtatcttctgcattctcgagaacccagatgtgtgtcttcttatacagaacct 1754
 DB 1020 TCGTTCAGCTTCTTCAATTTTGGACAAACCGAGTGTGGCTTCTTATATACATGACCT 1079
 QY 1755 gatgaacgaagaagatgcatagtcagaa 1784
 DB 1080 GATGAACCGCAAGATGATGCTCGAAT 1109

RESULT 10
 AF129003
 LOCUS 2099 bp mRNA linear PLN 22-DEC-1999
 DEFINITION Nicotiana tabacum diacylglycerol acyl-CoA acyltransferase (DAGAT)
 ACCESSION AF129003
 VERSION AF129003.1 GI:6625652
 KEYWORDS
 SOURCE
 ORGANISM
 Nicotiana tabacum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
 Bouvier-Nave,P., Benveniste,P., Oelkers,P., Sturley,S.L., and
 Schaller,H.
 Expression in yeast and tobacco of plant cDNAs encoding acyl
 CoA:diacylglycerol acyltransferase
 Eur. J. Biochem. 267 (1), 85-96 (2000)
 2 (bases 1 to 2099)
 Benveniste,P.
 Direct Submission
 Submitted (17-FEB-1999) Plant Molecular Biology Institute,
 C.N.R.S., 28 rue Goethe, Strasbourg 67083, France
 Location/Qualifiers
 1..2099
 /organism="Nicotiana tabacum"
 /cultivar="xanthi SH6"
 /db_xref="taxon:4091"
 /issue_type="call derived from leaf protoplasts"
 1..2099
 /gene="DAGAT"
 208..1806
 /gene="DAGAT"
 /function="catalyzes the acylation of diacylglycerol by
 oleoyl-CoA to give triacylglycerol"
 /note="similar to acyl-CoA cholesterol acyltransferase"
 /codon_start=1
 /product="diacylglycerol acyl-CoA acyltransferase"
 /protein_id="AAPI945.1"
 /db_xref="GI:6625653"

translation="MYIMLPESVEMTTTTTSGIENLNSDLNHSVRRRSGNGPEAA
 SATNSDAMSDDRNDVCGSAGLFTVVERSKSVGESDVIKEDRDNDVANGESK
 SETTTTPKFAVRAAPAHRIKESPSIDAIFKOSHAGPLNCVYVLIIVNSLIT
 ENIMYGLIRAGFESSKSLDWPFLMCCSIOTLPLAATLVERLADQRLTERAVY
 TLIITTPAAILYPIVLITGDSAPLRVLIIMLVACTYMKLIVASHNHDMLANS
 TPKDSGSDSYSDSPKSLATPMVAPLCTOLSPHTPCIRKVMVAKQFTIKVIFG
 LMGTLEOINIVYNSOHPKNGNLVIAIEVLIKSVNLIYVLMCMFECFHMNLIL
 AEILCFGDEEFTKDMNAKTIDEVWRMNMVHVMVHIFPCPLRNGIPGVAILLA
 FLVSAVFEHLIAVPCRLFKWMAFWGIMFQVPLVILTNFLQNKFOSSVYGMNMFCEFF
 CLTSGPVCVLLIYHDVNNRKSAR"

BASE COUNT

543 a 405 c 485 g 666 t

OY	666	lcaatcagaaatcttatgaaatgtagatgyltgltgatcaagaagattctggttga	722
Db	535	TAATATATTGAAATTTAATGAAGTATGGCTGGCTGATCAATATCAGAAATTTGGTTTAGTT	594
OY	728	caaatcgctcgagagattggcgcgtttccatggtgttataccctttgatcttcc	787
Db	595	CAACATCGCTTAAGGATTTGGCCACGCTAAAGTGTGTAGCTTCAGATTTTGTTCAC	654
OY	788	tgagctgctttagcttgtagaagattgtagtactcagaatacatatacgaaccgttgc	847
Db	655	TGCGTTCAATTTCTGTGCGAGAAATGGTGTGAATTAATTTAACTGATCGATGGGTGCAG	714
OY	848	tctttctcatatattatataccctgacagaggtttgtatccagtttacgtacc	907
Db	715	TCCTTCTTCATGTTTACATCAACAACATGGAAATCTTTGTTCCAGTTGTTGTTCATTTTA	774
OY	908	gggtgattctcgcttttttatacaggttgcctttgatgtctctacttgcattgtgtgc	967
Db	775	GGTGTGANTCGCTGTTCTTCAAGGTGTCAGCTAAAGCTCTTGGCTTCCACTGTATGTGT	834
OY	968	taaatgtgattcttattatgctacatacagatagatataagatccctcagcaatgtagctg	1022
Db	835	TGAAGCTCGTTTCTCTACGCAACATCAACTATGATTGAGTACTTGCAAATATCACTTG	894
OY	1028	ataag-----ggcaatccagaagctctcctaactaagttga	1066
Db	895	ATAGGTGGGAAGCTATGTCAGAGTACTGSAACCTGTGACCTGATATATGTATAGCTTTA	954
OY	1067	agagcttgacatattcaatggtcgctcccaatgtgttatcaagcaagtatccactt	1126
Db	955	AGAGTGTGCATATCTCAATGAGTGTGCTCTACATTTGTGTTACAGCAACGATACCTCGGA	1014
OY	1127	ctgcatgltatacgaagaggttggtggtcgtcgtcaattgcaaaactggtcatataccg	1186
Db	1015	CAGCTGTGATTCGGAAGGTTGGGTGGTAAGCAACTAATTAACCTGGTAATATTCACAG	1074
OY	1187	gattcatggatttataataaacaatatataaactcctatgttcaaggaaactcaagatc	1246
Db	1075	GACTCATGGGANTTTATTTAAACGATGATCAATTAACCCGATCGTTCAAAATTCATCAACT	1134
OY	1247	cttgaaggagacattctatatagtctatgaaagagtglttgaagcttgcagttccaaatc	1306
Db	1135	CTCTGAAGAAGACCTTTTATATATCCATTTGAGAGGGCTTGGAAGCTTTCTGTTCCAAAT	1194
OY	1307	tatatgtgtcctcgtcatgtcttaccgcttcttcaaccttggttaaacatatgtgcag	1366
Db	1195	TATATGTGTGCTGTGCATGTTTATTTGTTTTCACACCTGTGGCAAAATATCTTGCTG	1254
OY	1367	agcttcctccttcoggagatcggtgaactctcaaaagattggttgaatgcaaaaggtgtg	1426
Db	1255	AACTTCTGTGCTTTGGGACCGGTAGGAGTTTATTAAGATTTGGTGGAAATGGCAGACAGTGG	1314
OY	1427	gaagattcggagaatgtggaatgacccgttgcataaagatggtgttgcacatatact	1486
Db	1315	AGGAGTACTGGAAATGTGGAAATATGCCGTGCATTAATAGTATGGTGGCAATATATACT	1374
OY	1487	lcccggttgctggcagcaagatataccaagaacctgcacatatacttgcttccagact	1546
Db	1375	GTGCATATGCTTCAAAATGCAATATCAAAATAGTGGCAGTTTGTATGCCGTTTCTGTGCT	1434
OY	1547	ctgagactcttcaatgagcatatgcatcgaggttccctgtcgctcttcaagctatgggctt	1606
Db	1435	GTGGCAGTTTTTATAGAGCTGTGCGTTGGCAGTCCCTGCCAAATATTCAGAGTTTGGGCGT	1494
OY	1607	tctcttgagattatgtttaagtgcccttggctcttcaatcaacaatatatacgaagaaggt	1666
Db	1495	TCGTGGGATATCATGCTTACAGGTTCTCTTCGTAAATCGTACATAATTTACTTGCAGAGAAAGT	1554
OY	1667	t---tgctcaacgggtgggagacatgactctgtgttcaatctcttgatcttgcgaac	1723
Db	1555	TCAAAACCTCAATGATGGGCAATATGATGTTCTGCTCTCTCTGTATCTTTGGTCAAC	1614
OY	1724	cgatgtgtgtctcttattataccaagactcgtatgaaccganaag	1768

Db	1615	CTATGTGTGTTCTACTACACACATTGATGATGAAAG	1659
RESULT 12	AC005917	92721 bp	DNA linear
LOCUS	AC005917		PLN 05-APR-2000
DEFINITION	Arabidopsis thaliana chromosome II section 113 of 255 of the complete sequence. Sequence from clones F27F23, F3P11.		
ACCESSION	AC005917	AB02093	
VERSION	AC005917.2	GI:6598497	
KEYWORDS	HTG.		
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana		
REFERENCE	Eukaryota: Viridiplantae; Streptophyta; Embryophyta: Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots; Rosidae: eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 92721)		
AUTHORS	Lin, X., Kaul, S., Rounsley, S.D., Shea, T.P., Benito, M.-I., Town, C.D., Fujii, C.Y., Mason, T.M., Bowman, C.L., Barnstead, M.E., Feldhuth, T.V., Buell, C.R., Ketchum, K.A., Lee, J.-J., Ronning, C.M., Ko, H., Moffat, K.S., Comin, L.A., Shen, M., VanAken, S.E., Umayam, L., Tallon, L.J., Gill, J.E., Adams, M.D., Carrera, A.J., Creasy, T.H., Goodman, H.M., Somerville, C.R., Copenhaver, G.P., Preuss, D., Nierman, W.C., White, O., Eisen, J.A., Salzberg, S.L., Fraser, C.M. and Venter, J.C.		
TITLE	Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana		
JOURNAL	Nature 402 (6763), 761-768 (1999)		
MEDLINE	20083487		
PubMed	10617197		
REFERENCE	2 (bases 1 to 92721)		
AUTHORS	Lin, X.		
TITLE	Direct Submission		
JOURNAL	Submitted (09-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA		
COMMENT	On Dec 17, 1999 this sequence version replaced gi:4191771. The sequence and annotation of chromosome 2 were merged from those of the individual clones on this chromosome after removing overlaps. For detailed information, please see the TIGR web site (http://www.tigr.org/tdb/at/at.html).		

Genes were identified by a combination of three methods: Gene prediction programs including GRATL (<http://arthur.epm.ornl.gov/pub/gratl>), GeneFinder (Phil Green, University of Washington), Genscan (Chris Burge, <http://genomic.stanford.edu/Genscan.html>), and NetPlantGene (<http://www.cbs.dtu.dk/services/NetGene2/>). Searches of the complete sequence against a peptide database and plant EST databases at TIGR, and manual curations based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named similarly but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats were identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Genes are numbered from the top to bottom of the chromosome.

We thank the CSHL/MashU/ABI consortium for sequencing BAC clones F2P23, F5J6, T17A5, and T1316, the ESSA group for sequencing clone F13Jd, and Scott Jackson, Jining Jiang, Klaus Meyer, Eric Richards and Satoshi Tabaat for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khaila, Michael E. Heaney, Lily Fu, Peng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

Department of Energy and the US Department of Agriculture.

Address all correspondence to: ateltgr.org.

FEATURES

source

Location/Qualifiers
1. 92721

/organism="Arabidopsis thaliana"

/cultivar="Colombia"

/db_xref="taxon:3702"

/chromosome="II"

misc_feature

mRNA

/note="Sequence from clone F27E23"
complement(join<1311..1511,1608..2357,2445..2873,
3102..3203,3293..3567,3957..4070,4242..4380,4652..4824,
5200..5422))

gene

/gene="At2g19410"

/note="F27E23.21: contains a protein kinase domain profile
(PDOC00100)"

CDS

/product="putative protein kinase"
/protein_id="AADI0143.1"

misc_feature

/note="Sequence from clone F27E23"

mRNA

complement(join<1311..1511,1608..2357,2445..2873,
3102..3203,3293..3567,3957..4070,4242..4380,4652..4824,
5200..5422))

gene

/gene="At2g19410"

CDS

/product="putative protein kinase"
/protein_id="AADI0143.1"

misc_feature

/note="Sequence from clone F27E23"

mRNA

complement(join<1311..1511,1608..2357,2445..2873,
3102..3203,3293..3567,3957..4070,4242..4380,4652..4824,
5200..5422))

gene

/gene="At2g19410"

CDS

/product="putative protein kinase"
/protein_id="AADI0143.1"

misc_feature

/note="Sequence from clone F27E23"

mRNA

complement(join<1311..1511,1608..2357,2445..2873,
3102..3203,3293..3567,3957..4070,4242..4380,4652..4824,
5200..5422))

gene

/gene="At2g19410"

CDS

/product="putative protein kinase"
/protein_id="AADI0143.1"

misc_feature

/note="Sequence from clone F27E23"

mRNA

complement(join<1311..1511,1608..2357,2445..2873,
3102..3203,3293..3567,3957..4070,4242..4380,4652..4824,
5200..5422))

gene

/gene="At2g19410"

CDS

/product="putative protein kinase"
/protein_id="AADI0143.1"

misc_feature

/note="Sequence from clone F27E23"

mRNA

complement(join<1311..1511,1608..2357,2445..2873,
3102..3203,3293..3567,3957..4070,4242..4380,4652..4824,
5200..5422))

gene

/gene="At2g19410"

CDS

/product="putative protein kinase"
/protein_id="AADI0143.1"

misc_feature

/note="Sequence from clone F27E23"

mRNA

complement(join<1311..1511,1608..2357,2445..2873,
3102..3203,3293..3567,3957..4070,4242..4380,4652..4824,
5200..5422))

gene

/gene="At2g19410"

CDS

/product="putative protein kinase"
/protein_id="AADI0143.1"

misc_feature

/note="Sequence from clone F27E23"

mRNA

complement(join<1311..1511,1608..2357,2445..2873,
3102..3203,3293..3567,3957..4070,4242..4380,4652..4824,
5200..5422))

gene

/gene="At2g19410"

CDS

/product="putative protein kinase"
/protein_id="AADI0143.1"

misc_feature

/note="Sequence from clone F27E23"

mRNA

complement(join<1311..1511,1608..2357,2445..2873,
3102..3203,3293..3567,3957..4070,4242..4380,4652..4824,
5200..5422))

gene

/gene="At2g19410"

CDS

/product="putative protein kinase"
/protein_id="AADI0143.1"

misc_feature

/note="Sequence from clone F27E23"

mRNA

complement(join<1311..1511,1608..2357,2445..2873,
3102..3203,3293..3567,3957..4070,4242..4380,4652..4824,
5200..5422))

gene

/gene="At2g19410"

CDS

/product="putative protein kinase"
/protein_id="AADI0143.1"

misc_feature

/note="Sequence from clone F27E23"

mRNA

complement(join<1311..1511,1608..2357,2445..2873,
3102..3203,3293..3567,3957..4070,4242..4380,4652..4824,
5200..5422))

gene

/gene="At2g19410"

CDS

/product="putative protein kinase"
/protein_id="AADI0143.1"

misc_feature

/note="Sequence from clone F27E23"

mRNA

complement(join<1311..1511,1608..2357,2445..2873,
3102..3203,3293..3567,3957..4070,4242..4380,4652..4824,
5200..5422))

gene

/gene="At2g19410"

CDS

/product="putative protein kinase"
/protein_id="AADI0143.1"

misc_feature

/note="Sequence from clone F27E23"

mRNA

complement(join<1311..1511,1608..2357,2445..2873,
3102..3203,3293..3567,3957..4070,4242..4380,4652..4824,
5200..5422))

gene

/gene="At2g19410"

CDS

/product="putative protein kinase"
/protein_id="AADI0143.1"

misc_feature

/note="Sequence from clone F27E23"

mRNA

complement(join<1311..1511,1608..2357,2445..2873,
3102..3203,3293..3567,3957..4070,4242..4380,4652..4824,
5200..5422))

gene

/gene="At2g19410"

CDS

/product="putative protein kinase"
/protein_id="AADI0143.1"

misc_feature

/note="Sequence from clone F27E23"

mRNA

complement(join<1311..1511,1608..2357,2445..2873,
3102..3203,3293..3567,3957..4070,4242..4380,4652..4824,
5200..5422))

121 cgtcaaacgcttctcgaatgacgatttgattctctgctgactactacggtgacgag 180
|||||
22828 CGTCAACGCTTTTCGAATGGCGATTTCGCTGCTGCTTACTACTAGCGACGAG 22887
|||||
181 aacggtgagcgagagatctctgactcttgataagctctgctgacggaatcagatcgat 240
|||||
22888 AACGGGCGGGAAGTTCGTCGATCTTGATAGGCTTCGTCGAGAAATCGAGATCGGAT 22947
|||||
241 tcttcaacgagactctctctctgctgctccgataaattctctctcgtgatagtttgga 300
|||||
22948 TCTTTCACGAGACTTCTCTCTCTGCTCGGATTAATTTCTCTTCGATGATGTTGGA 23007
|||||
301 gctccgcgcgaagtaggagatcgatattccctgtgttaacgagatcgctcaggagaca 360
|||||
23008 GCTCCCGCGGAGCTTAGGATCGGATTCGATTCGCTGTTAAACGATGACGCTCAGGAGACA 23067
|||||
361 gccaatctggcggagataataacggtggtgagataaagcggtgtggaagaagcggc 420
|||||
23068 GCCAATTTGGCGGAGATATATAGGCTGCGGATTAATACGCTGTGGAAGAGCGCGC 23127
|||||
421 ggagaaagagagagaaacgcgacgtacgtttacgtlaccgacgctcgttccagctcat 480
|||||
23128 GGAGAAAGAGAGAGAAACCGCGATGCTAGCTTACGTATGACCGCTCGGTTCCAGCTCAT 23187
|||||
481 cggagagcgagagagagatccacttagctccgacgcaactcacaacagagcagtcgcgga 540
|||||
23188 CGGAGGCGGAGAGAGAGTCCACTTACCTCCGACGCAATCTTCAACAGGCTTTAAATCTC 23247
|||||
541 ttatcaaacctctgtagttagtcttctctgctgataacgagtaactcaccagaaat 600
|||||
23248 ACAAACTCTTCGAATTTGGTGTTCCTGTTTATGGAATTTAGATTGGTGTGATTTG 23307
|||||
601 ctt---atgaagagcagatcgagattatcaaacctctgtgtagtagtcttctatctgta 657
|||||
23308 TTTGCATTCGACAGCCATCCCGATTATTCACACTCTGTGTAGTCTTATGCTGTA 23367
|||||
658 aacagtagactcatcatcgaataactctatgaagtagtgcgtgt 700
|||||
23368 AACAGTAGACTCATCATCGAAAATCTTATGAAGGTTGCTGTT 23410
|||||
RESULT 13
AF221132 1497 bp mRNA linear INV 14-JUL-2000
LOCUS Caenorhabditis elegans diacylglycerol acyltransferase mRNA,
DEFINITION complete cds.
ACCESSION AF221132
VERSION AF221132.1 GI:9049537
KEYWORDS
SOURCE
ORGANISM
Caenorhabditis elegans.
Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilita;
Rhabdilitida; Rhabdilitidae; Pelodetidae; Caenorhabditis.
REFERENCE
AUTHORS
Bouvier-Nave, P., Benveniste, P., Noirel, A. and Schaller, H.
TITLE
Expression in yeast of a cDNA from Caenorhabditis elegans encoding
a diacylglycerol acyltransferase (DGAT)
JOURNAL
REFERENCE
AUTHORS
Bouvier-Nave, P. and Benveniste, P.
TITLE
Direct Submission
JOURNAL
Submitted (05-JAN-2000) Plant Molecular Biology Institute, CNRS, 28
rue Goethe, Strasbourg 67083, France
FEATURES
Source
1..1497
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/db_xref="dbEST:C47874"
/db_xref="taxon:6239"
/note="Original clone YK453a2 provided by Yuji Kohara"
1..1497
/EC_number="2.3.1.20"
/function="terminal enzyme during triacylglycerol

biosynthesis"
/note="DGAT: functional diacylglycerol acyltransferase
when expressed in yeast"
/codon_start=1
/product="diacylglycerol acyltransferase"
/protein_id="AAF82410.1"
/db_xref="GI:9049538"
/translation="MOMROGRRRRRRPSETSNGSLASSRSPFNONGNSRKSSEMR
GCEKRVHTADSLSTSSGWTNFKFENLSTLLVLSNGRVALENVKYLITPLQ
WSTVEHNTSISWPNLALILCSNQLISVGMKILIERGLNGFAVFTSLVIA
HLTIPEVVLITHRKNRNPMSVVMGVYVLEAKFTSYGHVNWADARKITELKTOV
TDLAKCTDKPQFMDLKDLSHOMQAAOYPAWLTLSNIYFPAWAPLICEFEPLRL
IKHFLIKRTVELILFSLIALVOOAVPYRNSMKPLISEMEYRSCERILKLAIPN
HLIWLFEYFEHSPNLNIAELRERADREFYDPMNAETIGFPMKSMNTPVIRAVRH
IYSPMRNRFSSMSAFPVYEFYSAFFHEIIVSPLKIRLRMSYTGMMGQIPLSITDK
VVRGRGNIILWLSLIVGQPLAATLMYGDWYILNFGVSAVONOTVGI"
BASE COUNT 402 a 311 c 308 g 476 t
ORIGIN
Query Match 7.8%; Score 154.2; DB 3; Length 1497;
Best Local Similarity 56.7%; Pred. No. 1.3e-28;
Matches 325; Conservative 0; Mismatches 243; Indels 5; Gaps 2;
1078 tatlcatgtgcgtccacatgtgtltaacgacagatattccagctctgcatgata 1137
|||||
772 TACTTCATGGCGACCAACATGTGCTCGAATTCGAATTTCCAGATTGTTGCCAATT 831
|||||
1138 cggaaaggtgtggtgtgctgctcgaatttgcacaaactgtgcatattcaccgagattcagga 1197
|||||
832 CGGAAGCACTTTTGTATTAAGAACCCGTGAGCTTATCTTTATCGTTTGTATAGCT 891
|||||
1198 ttataataaacaataataataactcattgtcgaaggaactcaagaatcctt---tga 1253
|||||
892 GCACCTGTTCAACAAATGGGTTGTCCGACTGTCGGAATAGATGAACCTTTAACTGA 951
|||||
1254 agcgagactctatagctatgaaagagtgtaagacttcaagttcacaattatagtg 1313
|||||
952 ATGGAATACCTTAGATGTT-TGGAACGACTCTTGAACCTTGAACATCACTCAT 1010
|||||
1314 gtgcgtcgtcagcttctactgcttctcacccttggtaaacatattgacagagcttct 1373
|||||
1011 CTGGCTTCTATTCTTCTACACATTTCTTCATTCATTTTGAACCTTGATGCCGAGCTGCT 1070
|||||
1374 ctgcttcggagatcgtgaattcttacaagaatgtgtgaatgtcgaagaatgttgagatga 1433
|||||
1071 TCGATTTCGCCATCGAATCTACAGAGACTTTTGAATGACAGAGCATAGCATATTT 1130
|||||
1434 ctggagaaatgtggaatagccgttcataaagatgtgtgacatatatacttccgtg 1493
|||||
1131 CTGGAATCATGTAACATCCAGTTCCAGATTGCTGTCCACATCTACAGTCCAAAT 1190
|||||
1494 ctgacgcagcaagaatacaagaacactgcacattatcattctgttctcagtcagct 1553
|||||
1191 GATGGGTAAACATTTCTCAAAATAGAGCCATTTCTGTTGTTCTTCTGTCGCGAATT 1250
|||||
1554 cttcatagatcagatcagatcgttctctgtctctcctcaagatcagatcagcttctg 1613
|||||
1251 CTTCATGATATTCGTGTTCTGTGCCATTAAAGATTTCGATGTGCTCTACTATAGG 1310
|||||
1614 gatatgttcaagtgctgtctgtcttcaac 1646
|||||
1311 AATGATGGAGCAAAATTCCTATTCATATATAC 1343
|||||
RESULT 14
CEH19N07/14
LOCUS CEH19N07 11966 bp DNA linear INV 11-DEC-2001
DEFINITION Caenorhabditis elegans cosmid H19N07, complete sequence.
ACCESSION Z92835
VERSION Z92835.1 GI:2414214
KEYWORDS
HTG.
SOURCE
Caenorhabditis elegans.

ORGANISM
Caenorhabditis elegans
Eukaryota; Metazoa; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE
1 (sites)
TITLE
Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium
JOURNAL
Science 282 (5396), 2012-2018 (1998)
MEDLINE
99069613
REMARK
The C. elegans Sequencing Consortium.
REFERENCE
2 (bases 1 to 11966)
AUTHORS
Dobson, R.
TITLE
Direct Submission
Submitted (09-MAR-1997) Nematode Sequencing Project, Sanger
Institute, Hinxton, Cambridge CB10 1SA, England and Department of
Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
jesse.sanger@wustl.edu
COMMENT
On Sep 20, 1997 this sequence version replaced gi:187712.
Coding sequences below are predicted from computer analysis, using
predictions from GeneFinder (P. Green, U. Washington), and other
available information.
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is not the entire insert of clone H19N07.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone H19N07 is at 1 in this sequence. The
true left end of clone C06H2 is at 11864 in this sequence. The
start of this sequence (1..1104) overlaps with the end of sequence
AL032618.
The end of this sequence (11864..11966) overlaps with the start of
sequence Z75526.
For a graphical representation of this sequence and its analysis
see: [http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=H19N07)
name=H19N07. This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
FEATURES
source
1..11966
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="v"
/clone="H19N07"
complement(join(794..1000,1327..2193,2252..2659,
3131..3252,3299..3416))
/gene="H19N07.1"
complement(join(794..1000,1327..2193,2252..2659,
3131..3252,3299..3416))
/note="Predicted using GeneFinder
contains similarity to Pfam domain: PF00009 (Elongation
factor Tu family). Score=372.3, E-value=1.6e-108, N=1
CDNA EST EMBL:Z14437 comes from this gene
CDNA EST EMBL:R02147 comes from this gene
CDNA EST EMBL:T00446 comes from this gene
CDNA EST YK118F10.3 comes from this gene
CDNA EST YK167A5.3 comes from this gene
CDNA EST YK118F10.5 comes from this gene
CDNA EST YK167A5.5 comes from this gene
CDNA EST YK340C12.3 comes from this gene
CDNA EST YK41347.3 comes from this gene
CDNA EST YK319H12.3 comes from this gene
CDNA EST YK319H5.5 comes from this gene
CDNA EST YK340C12.5 comes from this gene
CDNA EST YK41347.5 comes from this gene
CDNA EST YK491H12.5 comes from this gene
CDNA EST YK653A7.3 comes from this gene

CDNA EST YK518D11.3 comes from this gene
CDNA EST YK518D11.5 comes from this gene
CDNA EST YK570G3.3 comes from this gene
CDNA EST YK570G3.5 comes from this gene
CDNA EST YK616C5.3 comes from this gene
CDNA EST YK616C5.5 comes from this gene
CDNA EST YK623B10.3 comes from this gene
CDNA EST YK623B10.5 comes from this gene
CDNA EST YK625D9.3 comes from this gene
CDNA EST YK649E6.3 comes from this gene
CDNA EST YK649E6.5 comes from this gene
CDNA EST YK653A7.5 comes from this gene
CDNA EST YK662A10.3 comes from this gene
CDNA EST YK662A10.5 comes from this gene
CDNA EST YK674H11.3 comes from this gene
CDNA EST YK674H11.5 comes from this gene
/codon_start=1
/protein_id="CAB07395.1"
/db_xref="GI:3878044"
/translation="MSGNNVASSFVPNANRPYPCGPTTPOEQAPEPTEDMEO
ADTSPAAYOPEVAPVAPVAPVATAPKKEPTEDLVAPLAKKQRTVYVD
DOTHEKHNMTVGHVADGKSTIGGLMFLMGDKRTLEKYEBAKKGREYVLSW
CMTDDEEREKGTIVEGRATFETKRRFTLLDARGHSPYPMIVAGNADLAVI
SARKEPFTGEGDQGTREHSLVTVAGVAKHLVILNKMDDPYTKHEERKELEGTL
TPFLKRIKGFNPETDITVYVCSGLTGAFLIKDRPTGSEGMVSGPCTEITDLPSYKR
DNGPVRCVAKYSEMGTVIIGKMSCGVKGDILVMPKQPOVQVLIWADVETE
RVAGDNTIKFKLKGILENDLOGFTICSPDSIAKGRVDEAVLEHRSITASGSC
VLIHISAVEYEVVKKILRPELGGCPIGTIGTSDWPRNGLOVAFHPCNCRGVI
ATIDKKTGKKRAKRVKDEKIMLESPEPVLREFEYLYLGFILRDGKTIATG
KVLKVE"
join(4390..4601,4656..4991,5134..5600,5767..6728,
6773..6890,6935..7727,7777..7882,7945..8190,8238..8310,
8420..8499)
/gene="H19N07.2b"
join(4390..4601,4656..4991,5134..5600,5767..6728,
6773..6890,6935..7727,7777..7882,7945..8190,8238..8310,
8420..8499)
/gene="H19N07.2a"
join(4390..4601,4656..4991,5134..5600,5767..6728,
6773..6890,6935..7727,7777..7882,7945..8190,8238..8310,
8420..8499)
/note="contains similarity to Pfam domain: PF00442
(ubiquitin carboxyl-terminal hydrolases family 2),
Score=55.1, E-value=5e-13, N=1; PF00443 (ubiquitin
carboxyl-terminal hydrolase family 2), Score=107.3,
E-value=9.6e-29, N=1; PF00917 (MATH domain), Score=69.4,
E-value=2.4e-18, N=1
CDNA EST YK179D8.5 comes from this gene
CDNA EST YK513H2.5 comes from this gene
/codon_start=1
/protein_id="CAB07397.1"
/db_xref="GI:3878046"
/translation="MOCSPDEDLITVPHDIPSYDESIDPFGEGHLSLIDIDCFKSF
MSRSDNRMKSPITIVRGITPRILLATCNOQGSHPHNSNRNENFGLCONDEL
LQKGMRCRGTVAVLEVLNADGPSIQKTHASHNTVEVDMGSNYOYPTCNKPGY
LVNDTILRCRFTADVPTGANYMWSDRHRTGIGLNQGTCTMNSITDSFTTTFR
RAYNMVGTETPNESNIVLMQRYVELMAVETNSITRAFMDKDAEAMQHVQ
EPCVILNDLETKMKKSGSEKSIPLNFRGMKSYICLDVDESSSTESRYDVALVL
CMSELEAFAYTTSFTLDDENKYDADGLOAEGKVVEPLPIHJOLMRFQVC
VEQKINRSPFPEPMNLASCELCPTDEVDYSLHAYVHGFEHGHVYVIVN
LHESAVOPTSSAKCKRDDVDVSTTDDAIVSNFGEKTMNSAYMLVVRNADID
DOSIIVDSIPIHETVWKHMFTAELOFVHDLRFKSAQKIDMFDSDDEAQRARDILR
RIKSKNFNRFLMRMTDSYSLERTOKLTSRSEPTIDCTDRILDTLISODETIVY
EFSNNIRPLCEYDAPLAEIVFYUUTMDKFTIIGHTFDCHEKFNYSRMKCMH
KMLNLFDEVEEYQSYNKKRPEEPQGVKRCIDDKKIFVAAEELGASTVSNKQ
LITRWGSGSRREPFDYMLNCKGLMTFRVYHDPRAKRRKRYOAIAMPIDISNHR
IOTKLEFQOLPMNGHELETLPELKGTVITINEKRTYIPVBSGSGKRLLOIGAPPL
SNQRYVOIYCENTLISLDQRTMYKLOALHCRLEVPIDELMSPGFEFLCPVHFIRE

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 04:59:09 ; Search time 5105.53 Seconds
(without alignments)
5247.537 Million cell updates/sec

Title: US-09-623-514a-23

Perfect score: 1985
Sequence: 1 attcctaagctctctctc.....ttatggtgaagcgagcgcg 1985

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 segs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	405.8	20.4	729	10	BI422326 EST532992
2	361	18.2	399	9	BE523720 M40H1STM
3	336.4	16.9	629	9	AA042298
4	320.4	16.1	545	9	AI992543
5	314.2	15.8	539	10	BI422212
6	300.4	15.1	855	12	AO958057
7	299.4	15.1	855	9	AM035727
8	249.6	12.6	561	9	AM035727
9	226.6	11.4	626	9	AV926912
10	220	11.1	663	12	AO958056
11	216	10.9	862	10	BE321213
12	199	10.0	261	10	BE525637
13	191.8	9.7	260	10	BE525720
14	187.6	9.5	515	9	AV925760
15	174.6	8.8	422	10	BF199515
16	166.2	8.4	572	10	BM309699
17	162.6	8.2	617	9	AV930897

18	146.4	7.4	682	10	BE247899	BE247899 NF038011D
19	146	7.4	763	10	BI308446	BI308446 EST529856
20	142	7.2	577	9	AM586836	AM586836 EST318459
21	139.6	7.0	565	10	BC078967	BC078967 H3034809-
22	138	7.0	1064	10	BM476315	BM476315 AGENCOURT
23	137.8	6.9	706	10	BM448355	BM448355 MEST289-G
24	136.8	6.9	528	9	AL381190	AL381190 MCB857D12
25	135	6.8	634	10	BF634363	BF634363 NF059D06D
26	132.4	6.7	746	12	BA466416	BA466416 BOHEK18TR
27	131.4	6.6	832	9	AW342534	AW342534 GLEHEST391
28	128.4	6.5	701	10	BM343936	BM343936 MEST130-B
29	127.4	6.4	700	10	BM341343	BM341343 MEST333-F
30	127	6.4	531	10	BE356202	BE356202 DGI_123_B
31	126.2	6.4	531	9	AA141461	AA141461 CK01784.5
32	123.4	6.2	1000	9	AF122129	AF122129 AF122129
33	122	6.1	267	10	BF632251	BF632251 NF040E03D
34	121.8	6.1	681	10	BM267747	BM267747 MEST371-E
35	121.8	6.1	682	10	BM075161	BM075161 MEST351-B
36	119.4	6.0	670	10	BM338485	BM338485 MEST227-F
37	119.4	6.0	672	10	BM334036	BM334036 MEST131-F
38	119.4	6.0	819	10	BI225365	BI225365 602950108
39	118.2	6.0	667	9	AU168524	AU168524 AU168524
40	117	5.9	857	9	AL520752	AL520752 AL520752
41	115.4	5.8	664	10	BM266971	BM266971 MEST388-E
42	114.8	5.8	655	9	AW775077	AW775077 EST334228
43	113.4	5.7	761	10	BG785713	BG785713 SEAU0C005
44	113.2	5.7	899	9	AL571777	AL571777 AL571777
45	113	5.7	648	10	BM336514	BM336514 MEST195-A

ALIGNMENTS

RESULT 1
BI422326
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source

BI422326 729 bp mRNA linear EST 16-AUG-2001
EST532992 tomato callus, TAMU Lycopersicon esculentum cDNA clone
cLEEC69N10 5' end, mRNA sequence.
BI422326
BI422326.1 GI:15196624
EST.
tomato.
Lycopersicon esculentum
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots:
Asteridae: euasterids I: Solanales: Solanaceae: Solanum:
Lycopersicon.
1 (bases 1 to 729)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.,
, Liang,F., Upton,J., Craven,M.B., Bowman,C.D., Ahn,S., Renning,
, C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato callus tissue
Unpublished (1999)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>.
Location/Qualifiers
1..729
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEEC69N10"
/clone_lib="tomato callus, TAMU"
/tissue_type="callus"
/dev_stage="25-40 days old"
/lab_host="XLI-Blue MRF"
/note="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
XhoI; Supplier: Giovannoni laboratory; cLEEC - Cotyledons
of seedlings 7-10 days post-germination were excised, cut
at both ends and placed on MS medium with no selection.
Mixed callus was harvested at 25 and 40 days and included

DEFINITION	24635 CD4-13 Arabidopsis thaliana cDNA clone E68277, mRNA sequence
ACCESSION	AA043298
VERSION	AA043298.1 GI:2414087
KEYWORDS	EST.
SOURCE	thale cress.
ORGANISM	Arabidopsis thaliana
REFERENCE	Eukaryote: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eucotsids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 629)
AUTHORS	Newman,T., deBull,J.,F.,J., Green,P., Keegstra,K., Kende,H., McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M., Retzel,E., and Somerville,C.
TITLE	Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones plant physiol. 106, 1241-1255 (1994)
JOURNAL	95148729
MEDLINE	On Sep 19, 1997 this sequence version replaced gi:1520456.
COMMENT	Contact: Thomas Newman MSU-DOE Plant Research Laboratory Michigan State University MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E. Lansing, MI Tel: 517-353-0854 Fax: 517-353-9168 Email: 22313c@pblm.cl.msu.edu Seq primer: T7.
FEATURES	Location/Qualifiers
SOURCE	1..629
	/organism="Arabidopsis thaliana"
	/strain="Columbia"
	/db_xref="taxon:3702"
	/clone="E6827"
	/clone_id="CD4-13"
	/tissue_type="seedling hypocotyl"
	/dev_stage="3 day-old"
	/note="Vector: pBluescript SK-; Site.1: EcoRI; Site.2: EcoRI; Using 5 ug of polyadenylated mRNA from 3 day-old Arabidopsis thaliana (Columbia) seedling hypocotyls as template and oligo d(T) as primer, first strand synthesis was catalyzed by Moloney murine leukemia virus reverse transcriptase (Pharmacia). Second-strand cDNA was made using the procedure of Gubler and Hoffman (1983) except that DNA ligase was omitted. After the second strand reaction, the ends of the cDNA were made blunt with Klenow fragment and EcoRI/NotI adapters (Pharmacia) were ligated to each end. The cDNA was purified from unligated adapters by spun-column chromatography using sephacryl s-300 and size-fractionated on a 1% low melting point mini-gel. Size selected cDNAs (0.5 - 1 kb) were removed from the gel using agarase (New England Biolabs), phenol:chloroform extracted and precipitated using 0.3 M NaOAc (pH 7)/ethanol. A portion of each cDNA size-fraction (0.1 ug) was co-precipitated with 1 ug of lambdaZapII (Stratagene/EcoRI digested), dephosphorylated arms and then ligated in a volume of 4 ul overnight. Each ligation mix was packaged in vitro using GigaPack II gold packaging extract (Stratagene). We have determined that although first strand cDNA synthesis was initiated using dT, almost all of the cDNAs begin 8-10 bp from the poly-A tail. The reason for the loss of the poly-A tail is most likely due to lower than anticipated nucleotide levels during the Klenow repair of ragged ends before the addition of linkers (3'-5' exo instead of 5'-3' pol). When this library is used please reference the ABRC and: Kleiber, J. et al. (1993) Cell 72:427-441."
BASE COUNT	150 a 142 c 164 g 157 t 16 others
ORIGIN	

QY	1128	tcacgtatatacgaaggggttggtggtcgtcgtcaatttgcgaacatggctatattcaccgg	1187			
Db	1	TCGATGATACGGAAGGTTGGTGCGCTCGTCATTTGGCAAACTGGTCATATTACCGG	60			
QY	1188	attaatggacttlaataagacaatataataalcctatctgtcaggaactcaagcatcc	1247			
Db	61	ATTCATGGGATTTTAAATAGAACAAATATATTAATCTATTGTGAGAACTCAAGCATATCC	120			
QY	1248	tttgaagagcgactcttctatatgctatttgaagagtggttgaagcttcagttccaaattt	1307			
Db	121	TTTGAAGCGCATCTTCATATATCTATTGAAAGAGTGTTGAAGCTTCAGTCCAAAATTT	180			
QY	1308	atattggtgagcttcgcacatgcttctactgcttccaccttggttgaacaatatggcaga	1367			
Db	181	ATATGTGTGGCTCTCGCATGTTCTACTGCTTTCTCCACCTTTGGTTAAACATATTGGCAGA	240			
QY	1368	gcttcctgccttcgggagtcgtgaattctacaagaattggtgagtcgcaaaagtgtggg	1427			
Db	241	GCTTCCTGCTTCGGGGAGTCGTGAATTTCTACAAAGATTTGGTGAATGCACAAAAGTGTGGC	300			
QY	1428	agattact-ggaagaatgtggaatatgctctgtctcataatgg--atggttcgactatata	1484			
Db	301	AGATTACGCGGAGAAATGTGGAATATGCTGTCCATTAATGGGATGGGATCGCATATATATA	360			
QY	1485	c--tttccgtgcttcgcgacgaacagataccaaagaactgcgc	1524			
Db	361	CCTTCCCGCTGCTTCGCGCACAGCATTTCCGCAAGACACGCC	402			
RESULT	4	A1992543	545 bp	mRNA	linear	EST 08-SEP-1999
LOCUS	A1992543/c	701558181	A. thaliana, Ohio State clone set	Arabidopsis thaliana		
DEFINITION		CDNA clone 701558181, mRNA sequence.				
ACCESSION	A1992543					
VERSION	A1992543.1	GI:5839448				
KEYWORDS		EST.				
SOURCE		thale cress.				
ORGANISM		Arabidopsis thaliana				
REFERENCE		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
AUTHORS		Chen, J., Momiyama, M., Chan, E., Mooney, M., Carroon, B., Gilliland, D., Wang, X., Hillman, J., Giegler, K., Kim, C., Doyle, M., Brzoska, P., Gorgone, C., Burns, D., Griffin, J., Mouanoutoua, M., Nguyen, D., Tan, R., Rose, M., Warren, B., Ton, B., Kastury, K., Borillo, C., Carpio, T., Policky, J., Suzuki, G., Argente, C., Shah, S., Nobriza, A., Murry, L., Turner, C., Krikorian, S., Elder, L. and Hanson, D.				
TITLE		Arabidopsis thaliana Gene Expression MicroArray				
JOURNAL		Unpublished (1999)				
COMMENT		Contact: David Smoller, Ph.D. Genome Systems, Inc., a wholly owned subsidiary of Incyte Pharmaceuticals, Inc. 4633 World Parkway Circle, St. Louis, MO 63134, USA Tel: 877-577-2733 Fax: 314-427-3324 Email: service@genomesystems.com.				
FEATURES		Location/Qualifiers				
source		1..545				
		/organism="Arabidopsis thaliana"				
		/db_xref="taxon:3702"				
		/clone="701558181"				
		/clone_jid="A. thaliana, Ohio State clone set"				
		/note="CDNA library was made from selected clones from the Arabidopsis thaliana Ohio State clone set."				
BASE COUNT		146 a 125 c 120 g 154 t				
ORIGIN						
Query Match		16.1%	Score 320.4	DB 9	Length 545	
Best Local Similarity		83.5%	Pred No. 3.6e-70			

QY	1128	tcacgtatatacgaaggggttggtggtcgtcgtcaatttgcgaacatggctatattcaccgg	1187			
Db	1	TCGATGATACGGAAGGTTGGTGCGTCTGCATTTTGCANNACTGGTCAATATTCACCGG	60			
QY	1188	attaatggagcttlaataagacaatataataatcctatctgtcaggaactcaagcatcc	1247			
Db	61	ATTCATGGGATTTTAAATAGAACAAATATATTAATCTATTGTGAGAACTCAAGCATATCC	120			
QY	1248	tttgaagagcgatctctatatgtctatttgaagagtggttgaagcttcagttccaaattt	1307			
Db	121	TTTGAAGCGCATCTTCTATATATCTATTGAAAGAGTGTTGAAAGCTTCAGTCCAAAATTT	180			
QY	1308	atattggtgagcttcgcattgtcttactgcttccacttccacttggtaacaatatggcaga	1367			
Db	181	ATATGTGTGGCTCTGCGATGTTCTACTGCTTTCTCCACTTTGGTTAAACATATTGGCAGA	240			
QY	1368	gcttcctgccttcgggagtcgtgaattctacaagaattggtgagtcgcaaaagtgtggg	1427			
Db	241	GCTTCTGTGCTTGGGGAGTCGTGAATTTCTACAAAGATTTGGTGAATGCACAAAAGTGTGGC	300			
QY	1428	agattact-ggaagaatgtggaatatgctctgtctataatgg--atggttcgactatata	1484			
Db	301	AGATTACGCGGAGAAATGTGGAATATGCTGTCCATTAATGGGATGGTCCGACATATATATA	360			
QY	1485	c--tttccgtgcttcgcgacgaacaagataccaaagaactgcgc	1524			
Db	361	CCTTCCCGCTGCTTCCGACACAAAGATTTCCCAAGACACACCC	402			
RESULT	4	A1992543	545 bp	mRNA	linear	EST 08-SEP-1999
LOCUS	A1992543/c	701558181	A. thaliana, Ohio State clone set	Arabidopsis thaliana		
DEFINITION		CDNA clone 701558181, mRNA sequence.				
ACCESSION	A1992543					
VERSION	A1992543.1	GI:5839448				
KEYWORDS		EST.				
SOURCE		thale cress.				
ORGANISM		Arabidopsis thaliana				
REFERENCE		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
AUTHORS		Chen, J., Momiyama, M., Chan, E., Mooney, M., Carroon, B., Gilliland, D., Wang, X., Hillman, J., Giegler, K., Kim, C., Doyle, M., Brzoska, P., Gorgone, C., Burns, D., Griffin, J., Mouanoutoua, M., Nguyen, D., Tan, R., Rose, M., Warren, B., Ton, B., Kastury, K., Borillo, C., Carpio, T., Policky, J., Suzuki, G., Argente, C., Shah, S., Nobrega, A., Murry, L., Turner, C., Krikorian, S., Elder, L. and Hanson, D.				
TITLE		Arabidopsis thaliana Gene Expression Microarray				
JOURNAL		Unpublished (1999)				
COMMENT		Contact: David Smoller, Ph.D. Genome Systems, Inc., a wholly owned subsidiary of Incyte Pharmaceuticals, Inc. 4633 World Parkway Circle, St. Louis, MO 63134, USA Tel: 877-577-2733 Fax: 314-427-3324 Email: service@genomesystems.com.				
FEATURES		Location/Qualifiers				
source		1..545				
		/organism="Arabidopsis thaliana"				
		/db_xref="taxon:3702"				
		/clone="701558181"				
		/clone_jid="A. thaliana, Ohio State clone set"				
		/note="CDNA library was made from selected clones from the Arabidopsis thaliana Ohio State clone set."				
BASE COUNT		146 a 125 c 120 g 154 t				
ORIGIN						
Query Match		16.1%	Score 320.4	DB 9	Length 545	
Best Local Similarity		83.5%	Pred No. 3.6e-70			

The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: atc@ligr.org

For additional information, see <http://www.ligr.org/tdb/at/at.html>
Seq primer: TR
Class: Shotgun.

FEATURES

source

Location/Qualifiers
1..605
/organism="Arabidopsis thaliana"
/strain="Landsberg erecta"
/db_xref="taxon:3702"
/clone="PERAU62"
/clone_1ib="LERA"
/note="Organ: Leaf; Vector: PHOS1; Total genomic DNA was sheared to 0.9-1 kbp before ligation."

BASE COUNT 188 a 122 c 90 g 205 t
ORIGIN

Query Match 15.1%; Score 300.4; DB 12; Length 605;
Best Local Similarity 99.7%; Pred. No. 4.2e-65;
Matches 301; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1677 gctgggaacatgatactctgctgctcattcctgacgaacagatgtgtgct 1736
|||||
Db 452 GGTGGGGAACATATCTTCGTGTCATCTTCATCTTCGACACACCATGTGTGT 393
QY 1737 tcttattaccagaccgtgtagaaccgaaagagatgcatgtaagaactgttcaaa 1796
|||||
Db 392 TCTTTTATACACGACCTGATGAAACCGAAAGATGATGATGAACACATGTTCAAA 333
QY 1797 aatgagcttctcacaactcctgagccctggtgagtcctccctgagtggtgtgttc 1856
|||||
Db 332 AAATGACCTTCTTCAACATCTATGCGCTGCTGATGCTGATGCTGATGCTGCT 273
QY 1857 tgaatgctaaacgacaatagtgcttaaacattgagaagaagaagaatagagtgct 1916
|||||
Db 272 TGATGCAAAACGACAAATATGTGTATACCATTTGAAGAAAGAAATTTGACTTGT 213
QY 1917 tgaatgctaaacgacaatagtgcttaaacattgagaagaagaagaatagagtgct 1976
|||||
Db 212 TGTATCTGCAAAATTTTGTAGAGACGACGCAACCGCTTGTGATTTGTATGCTGTA 153
QY 1977 ag 1978
||
Db 152 AG 151

RESULT 7
AM349274/c 855 bp mRNA linear EST 04-OCT-2000
LOCUS
DEFINITION GM210004B21H12 Gm-r1021 Glycine max cDNA clone Gm-r1021-1536 3',
mRNA sequence.
ACCESSION AM349274
VERSION AM349274.1 GI:6846984
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseolae;
Glycine.

REFERENCE 1 (bases 1 to 855)

AUTHORS Vodka, L., Kelm, P., Shoemaker, R., Retzel, E., Khanna, A., Coryell, V.,
Erpedling, D., Kaph, C., Snoop, E., Pardinas, J., Liu, L. and Lewin, H.
TITLE A Functional Genomics Program for Soybean (NSF 9872565)
JOURNAL Unpublished (1999)
COMMENT Other ESTs: A1441040
Contact: Vodka, L.O., PI, A Functional Genomics Program for
Soybean (NSF 9872565)
Lewin, H. A., Director, Keck Center for Comparative and Functional

Genomics
University of Illinois
Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582
Email: l-vodkin@uiuc.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134. For further information
call: (800) 430-0030 or (314) 427-3322 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or info@genome
systems.com web site: www.genomesystems.com
Seq primer: 5'-TTTTTTTTTTTTTTTTT(A/C/G)-3'.

FEATURES

source

Location/Qualifiers
1..855
/organism="Glycine max"
/cultivar="Williams"
/db_xref="taxon:3847"
/clone="Gm-r1021-1536"
/clone_1ib="Gm-r1021"
/tissue-type="root"
/lab_host="XL10-Gold"
/note="Vector: pluscript II XR; Site: 1; EcoRI; Site: 2;
XhoI; Library Gm-r1021 is a sequence-driven, rereaked set
of the original library Gm-cl004 which was prepared from
root cDNA. The mRNA was isolated from entire roots of 8
day old 'Williams' seedlings which were propagated on
paper towels with distilled water. Stratagene's cDNA
Synthesis Kit (catalog #200401) was used to synthesize
the cDNA. The Gm-cl004 library was constructed by Dr.
Paul Kelm & Virginia H. Coryell, Department of Biology,
Box 5640, Northern Arizona University, Flagstaff, AZ 86011
email: paul.kelem@uau.edu, virginia.coryell@uau.edu. The
contig analysis to select unique genes was performed by
the laboratory of Ernest Retzel, Computational Biology
Centers, University of Minnesota,
http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html
Rereacking was performed by Genome Systems, St. Louis,
http://www.genomesystems.com, and sequencing by the Keck
Center for Comparative and Functional Genomics,
University of Illinois,
http://www.life.uiuc.edu/biotech/keck.html."

BASE COUNT 287 a 170 c 153 g 214 t 31 others
ORIGIN

Query Match 15.1%; Score 299.4; DB 9; Length 855;
Best Local Similarity 71.8%; Pred. No. 8.5e-65;
Matches 382; Conservative 0; Mismatches 147; Indels 3; Gaps 1;

QY 1262 tctatagctatgtaagagtgtagaagcttcagttccaaattatagtgagctct 1321
|||||
Db 844 TNNNTTACGCCACNNNNNAGATNNNGAANNNNNGTTCCAAATTTATATGNGNCCT 785
QY 1322 gcatgttctactgcttcttccacacttggttaacatatggcagagcttctgcttcg 1381
|||||
Db 784 GCATGTTCTATGCTTTTCCNNNNNNNGNTAAATATCTCGCAACGCTTTCGATTTG 725
QY 1382 gggatcgtaaatctcacaagaatgtagaagtcacaagaatgtagaagtcacga 1441
|||||
Db 724 GTGATCGTGAATTTCTACAAAGATTGTGGAATGCCAAAACATCTGCAAGATTATTTGAAGA 665
QY 1442 tgtggaatagcctgttccataaagtagtgtagacatatatactccgctgtgcga 1501
|||||
Db 664 TGTGGAATATGCTGTTCCAAATGATGATGATGATGATGATGATGATGATGATGATG 605
QY 1502 gcaaatatcacaagaagcactgcacattatcattgcttctcgtcgtcgtcgtcgt 1561
|||||
Db 604 ACGGTCTACCAAGAGCTGCTGCTTTTAATTGNNNNNCTGGTTTCTGTTTATTCATG 545
QY 1562 agcatgacagcagcttctgcttctcctcaagcactatggagcttcttggaattatgt 1621
|||||
Db 544 ACGTGTGATGCTGCTGCTTCCCTTCCACATATTCAGATTGTGGGCTTTCGATGATATGT 485

QY	1622	ctcaagtgcccttggctcttcacacaaactacacacaggaagaagtttg--ctcaagc	1678
Db	484	TTGAGTTCCTTGGCTTGATCACTAATATTCGCAGAAATTAATTCAGAAACTCAATGG	425
QY	1679	ctggagaaatgaatctctcggatcattctctgatttcggaaacacgaatgttggcttc	1738
Db	424	TTGGAAATATGATTTTTTGCTTCATATTCAGATCTCTGGTCAGACCTATGTGTGACTGC	365
QY	1739	tttatatacagactcgtatgaaacgaaagaatcgatgltcatgaaacaactg	1790
Db	364	TATATCTACCATGACTTGATGAAATAGGAAGGCAAACTTGACTGGAAGCTACGG	313

RESULT	8
AM035727	
LOCUS	
DEFINITION	561 bp mRNA linear EST 18-MAY-2001
	EST281881 tomato callus, TAMU Lycopersicon esculentum cDNA clone
	clEC36M23, mRNA sequence.

ORGANIC	tomato.
SOURCE	EST.
KEYWORDS	AW035727.1
VERSION	GI:5894483
ACCESSION	AW035727

ORGANISM
Lycopodium obscurum
Eukarya: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta
Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots:
Asterales, euasterids I; Solanales: Solanaceae: Solanum;
Lycopersicon.

REFERENCE	1 (bases 1 to 561)
AUTHORS	Alcala,J., Vreblov,J., White,R., Matern,A.L., Vision,T., Holt,I.E

Liang, F., Upton, J., Craven, M.B., Bowman, C.L., Ahn, S., Roning, C.M., Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J. Generation of ESTs from tomato callus tissue. Unpublished (1999)

COMMENT

Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.

FEATURES	Location/Qualifiers
source	1. .561

```

/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLEC36M23"
/clone_1lb="tomato callus, TAMU"
/tissue_type="callus"
/dev_stage="25-40 days old"
/lab_host="XLI-Blue MRF"
/node="Vector: pBluescript SK(-); Site.1: EcoRI; Site.2:
XhoI; supplier: Giovannoni laboratory; clec - Cotyledons
of seedlings 7-10 days post-germination were collected, cut
at both ends and placed on MS medium with no selection.
Mixed callus was harvested at 25 and 40 days and included
undifferentiated masses. Tomato Callus EST Library"
BASE COUNT      138 a      111 c      118 g      194 t
          BRITAIN

```

Query Match	12.6%	Score 249.6	DB 9	Length 561
Best Local Similarity	67.7%	Pred. No. 2.9e-52		
Matches 401; Conservative	0	Mismatches 149;	Indels 42;	Gaps 2

OY 746 ggcgcgttccatggttgtaataccttcgaattcccttgctgccttaagltg 805
||| ||| | ||| ||| | ||| ||| ||| ||| ||| ||| |||
Db 3 ggccactttgatgtagctgctttagtcttccgatttcccctttccctggcttttctgtcg 62

Dy 806 agaattgtactcagaatatacatagaaacctgtgcatctttcttatatta 865
||||| ||| || | |||| ||| || | ||||| || ||||
Db 63 AGAAATGCGCAGAGAAGATATATGCATGAACATGTACTCTTTCACATAITTA 122

866 taccatgacagaggttttgtatccagtttaagtcacccaatggtgtatctgctttt 925

D	123	TAAGCAGAGCTTCCATTTTGTATCCAGTTCTGCTATCCTCAGGCTGTGATTTCTGCTTTC	182
Q	926	tatacagtgatcactttagatgctccctcaacttgcatgtggtctaaagtgtgttctatg	985
D	183	TATGGGGTGTACACATGTATGATGTTTGGTTCATGTGTGATGAACACTAGTTCTTATG	242
Q	986	ctctacactagctatgacataagatgccaggc-----cattcagctagtataggcca	1036
D	243	CACATACAAATTATGATATGAGACAGCTTGCACAACTCTGTGATATAGGGTGTGAATTCGG	302
Q	1037	atccctgaagctctcctactacgtctagctgtgaagaagcttggcatattcatgtgcgtccca	1096
D	303	AAATCAACTACTCTTACAAATGTTAGTTTCAGAGTTTGGCTACTTCATAGTTGCTCCAA	362
Q	1097	catgtgttatccagccaagttatccagcttctgcattacatacgyaaggtgtggtgtgtc	1156
D	363	CTTTATGCTATAC-----AGGGTGTGGCTGGCCCC	389
Q	1157	gtcaattgtgcaaaactgtgcatactaccgcgattcatgagatcttaataagacaatata	1216
D	390	GCCAACTCATCAAGCTGCGTAAATTTTACAGGATTATATGGGATTTTTCATTGAGCGATATA	449
Q	1217	taaatcctatgtgtcagaagactcaagaacatcccttlyaaagcgatctctcatatgtcatly	1276
D	450	TTAACCCCATTTGTGCGAAGCTCACGACACATTCATTGGAAGAAACCTTTATACGCCATCG	509
Q	1277	aaagagtgtaaaagcttccatctccaattatatagttgtgtgtgcattgtt	1328
D	510	AGAGGGATTATAGCTTTCAGTTCACAAATTTTATATGTCTGGCTCTGCAGATT	561

RESULT	9			
AV926912				
LOCUS	626 bp	mRNA	linear	EST 18-JAN-2002
DEFINITION	AV926912 K. Sato unpublished cDNA library, cv. Hatuna Nijo second			

leaf stage seedling leaves *Hordeum vulgare* subsp. *vulgare* cDNA
clone bascd36a20 5', mRNA sequence.

ACCESSION	AV926912
VERSION	AV926912.1
GI	GI:18222709

KEYWORDS	EST.
SOURCE	Hordeum vulgare subsp. vulgare.

ORGANISM
Hordaeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Strep-
tomonadophyta; Mameelioributa; T

REFERENCE 1 (bases 1 to 636)
; Triticaceae; Hordeum.
Spermatophyta; Magnoliophyta;
Triticaceae; Hordeum.

REFERENCE	1 (bases 1 to 620)
AUTHORS	Sato, K., Saisho, D. and Takeda, K.
TITLE	Barlow EST sequencing protocol

FILE
JOURNAL
COMMENT

CONTACT: JAGGASU SHIN 1
Center For Genetic Resource Inf
National Institute of Genetics

National Institute of Genetics
1111 Yata, Mishima, Shizuoka 41
Tel.: 81-559-81-6856

Tel: 01 559 01 0050
 Fax: 81-559-81-6855
 Email: tshini@renes.nic.ac.za

```

NAME:  configgenes.mrg.ac.jp.
FEATURES
source      1      626
Location/Qualifiers

```

```

/organism="Hordeum vul
/cultivar="Haruna Nitc

```

```

/culture="Indiana NJ"
/db_xref="taxon:112509"
/cclone="hasd26a20"

```

```

/clone_lib="K. Sato un
Nito second leaf stage

```

```

/dev stage="second year"
/tissue_type="seedling"

```

BASE COUNT	151 a	122 c	133 g
ORIGIN			

Query Match	11.48;	Score 226
Best Local Similarity	62.08;	Pred. No.

Matches 388; Conservative 0; Mismatch

Oy	619	ggatattcaacccctggtgtagtaeettcttcttgtaaacgaagtagcctaactcgaa	678
Dd	1	GGCCTTTTCATCTGTGCATTGTTTCTTGTTGGTAGAGTAACGGCAGCGCTATTATTAG	60
Oy	679	aacctatgaagaatgagtgtgtgatccaagaatcttcgtttagttcaaatgcagctg	738
Dd	61	AACCTAAATGAAGTATGGCTTATTATAAGCAACTGGCTTTTGTTCAACTAGAACATTTG	120
Oy	739	cgaagttagccgctttctatgttgtataacccttgaaccttcccttgctgcgccctt	798
Dd	121	CGGGCTATGCCCATCTTATGTGTGCTTGTAGTCCACAGCGCTTCCCTTGTGTCAATT	180
Oy	799	acggttggaaaattgtaactccaagaatacacalacaagaacctgtcatccttccat	858
Dd	181	TCAGTTGAAGAAGTTGGCGCTTCGTAAATTTATTACTGATGCTGTGGTACCCTGTTTCAT	240
Oy	859	attatlatcacatgacagaagglttgatcccagittaacgtoaccaatgaagtgatctc	918
Dd	241	ATCATTTCTTACAMCAMCAAGTAATGTATAATCCAGTGGCTTGATTCATTCTAATGTCAATTCCT	300
Oy	919	gctttttatcaagtgtaactttgatatgtctccacttgatcattggtgtgactaaagtgtt	978
Dd	301	GCGGTTGTGTCTGGTTTTTGTGTATGTTTATTGCTTGCATTGTTTGGCTGAAGCTCGTA	360
Oy	979	tcttatgtctactagctatgaaceaagaatccccagaccaaatgcagctgataagccaat	1038
Dd	361	TCTTTTGGCCATACAAACCATGACGTAAAGGCAATTAAACCATTAAGTGTAAGAAGTTGAT	420
Oy	1039	ccgtgaagtcctcactacgttagcttgaa-----gagcttgca	1077
Dd	421	TATGCAACCCACACAGATGACATGGATTAATTACAAGCTCCAACTTTAGGAGTCTAATG	480
Oy	1078	tattccatcagtcogctcccacattggtatataagccaagtatcacagytctgcagtata	1137
Dd	481	TACTTCATGATGCTGCCAACACTGTGCTATACGCCAACCTATCTCGAAGTCCGATGTT	540
Oy	1138	cggaaaggctbggtgctcgtlocaatttgcanaaatgtcatatalcaaccgaltcatgga	1197
Dd	541	AGAAAGGGTGTGGCTGATTTCGTCAAATTTATTCGTACTTGATATTTAACAGATATTCAAGC	600
Oy	1198	tttataataagaacatatataaac	1223
Dd	601	TTCAATTATTGACAGTACATAAATCC	626

RESULT	10
AQ958056	
LOCUS	
DEFINITION	AQ958056 663 bp DNA linear GSS-28-JAN-2000 LENAU627F LENA Arabidopsis thaliana genomic clone LENAU62, DNA sequence.
ACCESSION	AQ958056
VERSION	AQ958056.1 GI:6785757
KEYWORDS	GSS.
SOURCE	thale cress.
ORGANISM	Arabidopsis thaliana
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 663)
AUTHORS	Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Ullrichbach,T., Feildlyum,T., Liang,F., Creasy,T. and Fraser,C.M.
TITLE	Genomic survey sequencing of Landsberg erecta ecotype of Arabidopsis thaliana and identification of sequence-based polymorphisms
JOURNAL	Unpublished (2000)
COMMENT	Contact: Xiaoying Lin

Unpublished (2000)
Contact: Xiaoying Lin
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: xtli@igf.org
For additional information, see <http://www.tlgr.org/tldb/at/at.html>

FEATURES	Seq primer: TF
Source	Class: Shotgun.
Location/Qualifiers	
1..663	
/organism="Arabidopsis thaliana"	
/strain="Landsberg erecta"	
/db_xref="taxon:3702"	
/clone="LERAH62"	
/clone_1lb="LERA"	
/note="Organ: Leaf; Vector: pHOS1; Total genomic DNA was sheared to 0.9-1 Kbp before ligation."	
BASE COUNT	
ORIGIN	
191 a	129 c
	123 g
	220 t

[illegible]

RESULT	11
BG321213	
LOCUS	BG321213 862 bp mRNA linear EST 27-FEB-2001
DEFINITION	Zm04_05908_R Zm04_AAFc_ECOFC_cold_stressed_maize_seedlings Zea mays
ACCESSION	CDNA clone Zm04_05908, mRNA sequence.
VERSION	BG321213
KEYWORDS	BG321213.1 GI:13150891
SOURCE	EST.
ORGANISM	Zea mays.
REFERENCE	Zea mays. Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoidae; Andropogoneae; Zea. 1 (bases 1 to 862)
AUTHORS	Singh,J.A., Wakui,K., Couroux,P., De Moors,A., Harris,L.J., Hattori,T., J.I., Queillet,T., Robert,L.S., Sproll,D. and Tinker,N.A. Expressed Sequence Tags from Cold-Stressed Maize Seedlings Unpublished (2001)
TITLE	Contact: Singh,J.A.
JOURNAL	Eastern Cereal and Oilseed Research Centre
COMMENT	Agriculture and Agri-food Canada 960 Carling Avenue, Bldg. 20, Ottawa, Ontario, K1A 0C6, Canada Tel.: (613) 759-1662 Fax: (613) 759-1701 Email: singhja@agr.ca.
FEATURES	location/qualifiers
Source	1..862

```

/organism="Zea mays"
/cultivar="C0328"
/db_xref="taxon:4577"
/clone_11b="Zm04.AAC.C.ECORC_cold_stressed_maize_seedlings"
/clone_11b="Zm04.AAC.C.ECORC_cold_stressed_maize_seedlings"
/tissue_type="Leaf, crown"
/notes="Vector: Bluescript SK-/XhoI-EcoRI; Site1: Eco RI;
Site 2: Xho I; Lower temperature 50 C / hour from 22 to
12oc; bring to 50 in 1 hour from 12oc. Leave at 50c 2 days
photoperiod 16 hours. Light intensity was 125 uE-1."

```

Library prepared by in vivo mass excision from amplified
library.
BASE COUNT 234 a 146 c 165 g 300 t 17 others
ORIGIN

Query Match 10.9%; Score 216; DB 10; Length 862;
Best Local Similarity 63.2%; Pred. No. 1.1e-43;
Matches 349; Conservative 7; Mismatches 175; Indels 21; Gaps 1;

Y 823 aaatcatacataagaacctgtgtcattctctcatatattatatacacaagaagatt 882
B 9 AAGCATCATGGTGAACATGTGGTATTCTACCATCATCATATTAACAACATCTGCAT 68
Y 883 ttgtatcagattacgtacacctaaagtggtatctgtctttttatcagtgaccttg 942
B 69 GTCATATCCAGTTGTTGTGAKBHTTAAGTGTGACTCAGCAGTACTATTCGATTTGTCTA 128
Y 943 atgctcctcaacttgatgtgtgtcctaaagtgtgttcttctatcctacactagctatgac 1002
B 129 ATGTTTCTTGGCAGCATCATGTGATGAAGCTTGCTCTTATGACATTAACAATTATGAT 188
Y 1003 ataagatccccaagccaatgacgtatgaag-----gcaatcct 1041
B 189 ATAAAGGTAATTTGTCATAAAGTACTGAAAGGGTCTGCATATGAATAATTATGATCCT 248
Y 1042 gaagctctactactacgttagctgaagagctgtgcatattatcattgtcgtccacatg 1101
B 249 BAGAATATGAAGAAGATCCAAACCTTTAAAGTGTAGTACTTACTGTTGGCCCAACACT 308
Y 1102 tttatcagccaagtattacacagttctgcagtatacaggaaggtgtgtgtgtcgtcaa 1161
B 309 TGTATCCAGCCCAACTTATCTCAAACTCATATGATTAGAAAGGGTGGTGACCCAGCAA 368
Y 1162 ttgcaaaacagctgcatattaccggtatcattgagattatataataaataataat 1221
B 369 CTCATTAAGTGGCTGTTTATTACAGGCTTTATGGCTTCATATATKACCAATATTAAC 428
Y 1222 cctattgtcagaactcaagaacatccttgaagagcactctctatagctatgaaga 1281
B 429 CCAATTGTTGAAGAAATTCACCAACATCCACTGAAAGGAAATTTTGAATGCBATAGAA 488
Y 1282 gttgtgaagcttcagttcacaattatattatgtgtgtcctcattgtacgtcttc 1341
B 489 GTCCTTAACACTCTCAGTGCACCAATATATATGATGCTTTGCAATGCTATTTT 548
Y 1342 caacttggtta 1353
B 549 CATTATGTTA 560

RESULT 12
BE525637
LOCUS 261 bp mRNA linear EST 19-MAR-2001
DEFINITION M63K06STM Arabidopsis developing seed Arabidopsis thaliana cDNA
ACCESSION BE525637
VERSION BE525637
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE 1 (bases 1 to 261)
AUTHORS White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de
Iliarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.
TITLE A new set of Arabidopsis expressed sequence tags from developing
seeds. The metabolic pathway from carbohydrates to seed oil
JOURNAL Plant Physiol. 124 (4), 1582-1594 (2000)
MEDLINE 20567808
COMMENT Contact: Benning, C
Dept. of Biochemistry & Molecular Biology

FEATURES

source
1..261
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="600034122R1"
/clone_11b="Arabidopsis developing seed"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/note="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoI"
Location/Qualifiers
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824
, USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Clones were originally prepared at Michigan State University.
Arabidopsis Biological Resource Center, The Ohio State University,
309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210
USA, FAX: 6142920603 TEL: 6142929371.

BASE COUNT 69 a 55 c 71 g 66 t
ORIGIN

Query Match 10.0%; Score 199; DB 10; Length 261;
Best Local Similarity 100.0%; Pred. No. 1.5e-39;
Matches 199; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 411 aagagcgcgagagaagaagaacgcagatcgttactgatacgacgctgcgt 470
B 1 AAGAGCGCGGAGAGAGAGAAAGCGCGATCTACGTTTACGTACGACCGTCCGT 60
Y 471 tccagctatcggagagagagagagagctccttgccacccgaattccaacag 530
B 61 TTCACCTATCGAGAGCGGAGAGAGATCCACTTACGCTCCGCAAACTTCAACAGAG 120
Y 531 ccatcgcgatattcaacctctgtgtagttcttattgtctgaacagtagactat 590
B 121 CCATCCGCGATATTCAACCTCTGTGTAGTGTCTTATGCTGTAAACAGTAGACTCAT 180
Y 591 catcgaaaattcttagaag 609
B 181 CATCGAAATCTTATGAG 199

RESULT 13
BE525720
LOCUS 260 bp mRNA linear EST 19-MAR-2001
DEFINITION M63K06STM Arabidopsis developing seed Arabidopsis thaliana cDNA
ACCESSION BE525720
VERSION BE525720
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE 1 (bases 1 to 260)
AUTHORS White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de
Iliarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.
TITLE A new set of Arabidopsis expressed sequence tags from developing
seeds. The metabolic pathway from carbohydrates to seed oil
JOURNAL Plant Physiol. 124 (4), 1582-1594 (2000)
MEDLINE 20567808
COMMENT Contact: Benning, C
Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824
, USA
Tel: 517 355 1609

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 07:23:23 ; Search time 150.64 Seconds
(Without alignments)
3236.740 Million cell updates/sec

Title: US-09-623-514A-23
Perfect score: 1985
Sequence: 1 attcttgccttccttc.....ttatgtgttaagcgagcgcg 1985

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/prodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/prodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/prodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/prodata/2/ina/PCTUS.COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	336.4	16.9	629	4	US-09-103-754A-3	Sequence 3, Appli
2	141.8	7.1	1650	4	US-09-103-754A-2	Sequence 2, Appli
3	141.2	7.1	1976	3	US-09-165-042-2	Sequence 2, Appli
4	57	2.9	4011	1	US-08-121-057-3	Sequence 3, Appli
5	57	2.9	4011	2	US-08-509-187D-3	Sequence 3, Appli
6	57	2.9	4011	2	US-09-121-396-3	Sequence 3, Appli
7	57	2.9	4011	5	PCT-US93-08704A-3	Sequence 3, Appli
8	57	2.9	4079	1	US-08-121-057-2	Sequence 2, Appli
9	57	2.9	4079	2	US-08-509-187D-2	Sequence 2, Appli
10	57	2.9	4079	5	US-09-121-396-2	Sequence 2, Appli
11	57	2.9	4079	5	PCT-US93-09704A-2	Sequence 2, Appli
12	51.6	2.6	7218	1	US-08-232-463-14	Sequence 14, Appli
13	50	2.5	2040	3	US-09-165-042-4	Sequence 4, Appli
14	43	2.2	18596	4	US-09-318-448-11	Sequence 11, Appli
15	35.6	1.8	248	4	US-09-007-005-32	Sequence 32, Appli
16	35.6	1.8	248	4	US-09-244-796-32	Sequence 32, Appli
17	35.6	1.8	277	4	US-09-007-005-3	Sequence 3, Appli
18	35.6	1.8	277	4	US-09-244-796-3	Sequence 3, Appli
19	35.6	1.8	2367	4	US-09-056-556-201	Sequence 201, Appli
20	35.6	1.8	4411529	4	US-09-103-840A-1	Sequence 1, Appli
21	34.8	1.8	906	3	US-08-765-856-3	Sequence 3, Appli
22	34.8	1.8	906	4	US-08-935-009A-3	Sequence 3, Appli
23	34.8	1.8	908	3	US-08-765-856-1	Sequence 1, Appli
24	34.8	1.8	908	4	US-08-935-009A-1	Sequence 1, Appli
25	34.4	1.7	53526	3	US-08-658-136-2	Sequence 2, Appli
26	34.4	1.7	53577	3	US-08-658-136-1	Sequence 1, Appli
27	33.8	1.7	361	4	US-09-018-584A-9	Sequence 9, Appli

C 28	33.8	1.7	2397	1	US-07-891-942G-11	Sequence 11, Appli
C 29	33.6	1.7	1294	4	US-09-312-038-4	Sequence 4, Appli
C 30	33.6	1.7	2289	4	US-09-312-038-3	Sequence 3, Appli
C 31	33.4	1.7	279	2	US-08-623-906A-3	Sequence 3, Appli
C 32	33.4	1.7	289	4	US-09-007-005-17	Sequence 17, Appli
C 33	33.4	1.7	289	4	US-09-244-796-17	Sequence 17, Appli
C 34	33.4	1.7	7218	1	US-08-232-463-14	Sequence 14, Appli
C 35	33	1.7	80595	4	US-09-078-294-3	Sequence 3, Appli
C 36	32.8	1.7	1298	3	US-08-948-705-3	Sequence 3, Appli
C 37	32.6	1.6	5785	3	US-08-480-640A-221	Sequence 221, App
C 38	32.6	1.6	5785	4	US-08-686-968C-221	Sequence 221, App
C 39	32.6	1.6	5785	4	US-08-488-237A-221	Sequence 221, App
C 40	32.4	1.6	370	2	US-08-332-766A-8	Sequence 8, Appli
C 41	32.4	1.6	5000	3	US-09-104-070-1	Sequence 1, Appli
C 42	32.4	1.6	80246	4	US-09-078-294-4	Sequence 4, Appli
C 43	32.2	1.6	446	2	US-08-332-766A-26	Sequence 26, Appli
C 44	32.2	1.6	597	2	US-08-332-766A-19	Sequence 19, Appli
C 45	32.2	1.6	1383	2	US-08-630-822A-82	Sequence 82, Appli

ALIGNMENTS

RESULT 1
US-09-103-754A-3
: Sequence 3, Application US/09103754A
: Patent No. 6344548
: GENERAL INFORMATION:
: APPLICANT: Faresse, Robert
: APPLICANT: Cases, Sylvaine
: APPLICANT: Smith, Steven
: APPLICANT: Erickson, Sandra
: TITLE OF INVENTION: Diacylglycerol O-acyliran
: TITLE OF INVENTION: sferase
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Bozicevic & Reed
: STREET: 285 Hamilton Avenue, Suite 200
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94301
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/103,754A
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Field, Bret E
: REGISTRATION NUMBER: 37,620
: REFERENCE/DOCKET NUMBER: 6510-105p
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650 327 3400
: TELEFAX: 650 327 3231
: TELEX:
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 629 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-09-103-754A-3
Query Match 16.9%; Score 336.4; DB 4; Length 629;


```

ADDRESSER: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/509,187D
FILING DATE: 31-JUL-1995
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Lamport Hammitte, Ann
REGISTRATION NUMBER: 34,858
REFERENCE/DOCKET NUMBER: DCI-033CPDV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 4011 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1397..3046
US-08-509-187D-3

```

Query Match 2.9%; Score 57; DB 2; Length 4011;
Best Local Similarity 46.6%; Pred. No. 3.2e-07;
Matches 257; Conservative 0; Mismatches 285; Indels 9; Gaps 2;

1078 tatttcacatgctgcctccacatgctgctacgccaagtatccacgcttcgcatgata 1137
1138 cggaagagtggtgctgcctacatgctgctacgccaagtatccacgcttcgcatgata 1197
2351 agatggggttatgctgcctacatgctgctacgccaagtatccacgcttcgcatgata 2410
1198 ttataatagaacaatatataatccatattgctcaggaactcaaacatcccttgaaagc 1257
2411 TACATCTTTGAAGAGCTTTGTCCTTGTTCGGAATATCAACAGAGCCCTTCACG 2470
1258 gatctctatatgctatgaagaagtgtgaagcttcgcttccaaattatatgtgtg 1317
2471 GCTCGGT---TCGCGCTATGTGTATTAACCTCACTCTCCAGAGTGTGCTGATTC 2527
1318 cctcagatgctacgctctccacatgctgcttaacacatattgcaagcttcctgc 1377
2528 TTCCTTACTTTTTCCTTTTGCCTTTTGCACTGCTGCTCAATGCCCTTGCTATGATGTTACG 2587
1378 ttccggagatgctgaattacaaagtgtggaatgcaaaaagtgtggaagtattg 1437
2588 TTTGGTGAAGAGATGTCTTATTAAGATGTGGTGAAGCTCCACGCTCAATCCAACTATTA 2647
1438 agaagtggaatatacctccttcaataaagtgtgctcgcataat-----actcccg 1491
2648 AGAAGCTGAATGTGTGCTGCTATGATGCTGCTATATTTACTATGCTTACAGAGCTTTCTC 2707
1492 tgcctgcagagaatataccaaagaacatcgcacatattgcttccctagctctga 1551
2708 TGGTTTTCCTCAAGAGATTAATCTGCTGCCATGTTGCTGCTTCTGATCTGCT 2767

```

1552 gttctcatgactatgcacatgctcgttcgtctccttcaagaataggttctt 1611
2768 GTAGTACAGGAATATGCTTGCTGCTGCTTTTATCCGCTGCTTCG 2827
1612 gggatcatgt 1622
2828 CTCCTCATGTT 2838

```

RESULT 6
US-09-121-396-3
Sequence 3, Application US/09121396
Patent No. 5968749
GENERAL INFORMATION:
APPLICANT: CHANG, TA-YUAN
TITLE OF INVENTION: ACYL COENZYME A: CHOLESTEROL
TITLE OF INVENTION: ACYLTRANSFERASE (ACAT)
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSER: LAHIVE & COCKFIELD
STREET: 28 STATE STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/121,396
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/08/509,187
FILING DATE: 07/31/95
ATTORNEY/AGENT INFORMATION:
NAME: LAMPORT HAMMITTE, ANN
REGISTRATION NUMBER: 34,858
REFERENCE/DOCKET NUMBER: DCI-033CPDV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-2700
TELEFAX: (617) 742-4214
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 4011 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-121-396-3

Query Match 2.9%; Score 57; DB 2; Length 4011;
Best Local Similarity 46.6%; Pred. No. 3.2e-07;
Matches 257; Conservative 0; Mismatches 285; Indels 9; Gaps 2;

1078 tatttcacatgctgcctccacatgctgctacgccaagtatccacgcttcgcatgata 1137
2291 TACTCTTATTTGCTTCCTTACCTTATCTACCGTACAGCTATCCAGAACTCCACTGTA 2350
1138 cggaagagtggtgctgcctacatgctgctacgccaagtatccacgcttcgcatgata 1197
2351 AGATGGGTTATGCTGCTATGAAGTTTGCAAGGCTTTGTTGCTTTTCTATGCTGAC 2410
1198 ttataatagaacaatatataatccatattgctcaggaactcaaacatcccttgaaagc 1257
2411 TACATCTTTGAAGAGCTTTGTCCTTGTTCGGAATATCAACAGAGCCCTTCACG 2470
1258 gatctctatatgctatgaagaagtgtgaagcttcgcttccaaattatatgtgtg 1317

Db 2471 GCTGCTGT---TCTGCTCCATGTATTTAACTCCATCTTGGCAGSTGTGCTGATTC 2527
QY 1318 ctctgcatgtctactgtctctccaccltggtaaacatltggcagagctctctgc 1377
Db 2528 TTCCCTTACTTTTGTGCTTTTGGCACTGCTGCAATCCCTTGTGAGATGTACGC 2587
QY 1378 ttctggagatctgtatctctcaagatltgttggaagcaaaagtgtggagatltctg 1437
Db 2588 TTTGGTGACAGAGATGTTCTAATAAGATTGGTGAACCTCCAGTCATACCTCACTTAT 2647
QY 1438 agaatgtggaatatgtcctgtctcaataatgtatgttcgacataat-----acttccg 1491
Db 2648 AGAACCCTGGAATGTGTGGTCCATGACTGGCTATATTAATGCTTACAGAGACTTTCTC 2707
QY 1492 tgcctgcgcaagaataacaaagacactgcacatcatcatgtctctcagctctgcga 1551
Db 2708 TGGTTTTCTCCAGAGATTTCAATCTGCTGCACTGTACCTGCTTGTGATCTGCT 2767
QY 1552 gtcttcatagcatgtcagctcctgtctgtctctcaagctatggcttctt 1611
Db 2768 GTACTACACGAATATGCTTGGCTTGGCTTGAAGCTTTTCTATCCCGTGTCTGCTG 2827
QY 1612 gggatlatgt 1622
Db 2828 CTCTTCATGTT 2838

RESULT 7
PCT-US93-09704A-3

Sequence 3, Application PC/TUS9309704A
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: ACYL COENZYME A: CHOLESTEROL
TITLE OF INVENTION: ACYLTRANSFERASE (ACAT)
NUMBER OF SEQUENCES: 9
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/09704A
FILING DATE: October 12, 1993
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. SER. NO. 959,950
FILING DATE: October 14, 1992
APPLICATION NUMBER: U.S. SER. NO. 121,057
FILING DATE: September 10, 1993
ATTORNEY/AGENT INFORMATION:
NAME: LAMPORT HAMMITTE, ANN.
REGISTRATION NUMBER: 34,858
REFERENCE/DOCKET NUMBER: DCI-033CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-2700
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 4011 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
PCT-US93-09704A-3

Query Match 2.9%; Score 57; DB 5; Length 4011;
Best Local Similarity 46.6%; Pred. No. 3.2e-07;
Matches 257; Conservative 0; Mismatches 285; Indels 9; Gaps 2;
QY 1078 tattatgtgtctccacatgtgtatgcgaagtatccacgcttcgatgtata 1137
|| ||| | | ||||| || | | | || ||||| | | | | |

Db 2291 TACTTCTATTTTGTCTTACCCCTTATCTACCGTGACAGCTATCCAGAGATCCACGTGA 2350
QY 1138 cggaaaggttgggtgtgtcctcgtcatttgcaaacactgtcatatccaggatltcagga 1197
Db 2351 ACATGGGGTTATGTGCTATGAAAGTTTGACAGGTCTTTGGTGTCTTTTCTATCTGATAC 2410
QY 1198 ttataatagaacaalataataatcctatltgcaagaaactcaaacatcctltgaaaggc 1257
Db 2411 TACATCTTTGAAAGGCTTTGTGCCCTTGTGCGAATATCAACACAGACCCCTTACGC 2470
QY 1258 gatctctatgtcatltgaaagagtltgaaagcttcaagtlccaatttataatgtg 1317
Db 2471 GCTGCTGT---TCTGCTCCATGTATTTAACTCCATCTTGGCAGTGTGCTGATTC 2527
QY 1318 ctctgcatgtctactgtctctccaccltggtaaacatltggcagagctctctgc 1377
Db 2528 TTCCCTTACTTTTGTGCTTTTGGCACTGCTGCAATCCCTTGTGAGATGTACGC 2587
QY 1378 ttctggagatctgtatctctcaagatltgttggaagcaaaagtgtggagatltctg 1437
Db 2588 TTTGGTGACAGAGATGTTCTAATAAGATTGGTGAACCTCCAGTCATACCTCACTTAT 2647
QY 1438 agaatgtggaatatgtcctgtctcaataatgtatgttcgacataat-----acttccg 1491
Db 2648 AGAACCCTGGAATGTGTGGTCCATGACTGGCTATATTAATGCTTACAGAGACTTTCTC 2707
QY 1492 tgcctgcgcaagaataacaaagacactgcacatcatcatgtctctcagctctgcga 1551
Db 2708 TGGTTTTCTCCAGAGATTTCAATCTGCTGCACTGTACCTGCTTGTGATCTGCT 2767
QY 1552 gtcttcatagcatgtcagctcctgtctgtctctcaagctatggcttctt 1611
Db 2768 GTACTACACGAATATGCTTGGCTTGGCTTGAAGCTTTTCTATCCCGTGTCTGCTG 2827
QY 1612 gggatlatgt 1622
Db 2828 CTCTTCATGTT 2838

RESULT 8

US-08-121-057-2
Sequence 2, Application US/08121057
Patent No. 5484727
GENERAL INFORMATION:
APPLICANT: CHANG, TA-YUAN
TITLE OF INVENTION: ACYL COENZYME A: CHOLESTEROL
TITLE OF INVENTION: ACYLTRANSFERASE (ACAT)
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 STATE STREET, SUITE 510
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/121,057
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: LAMPORT HAMMITTE, ANN.
REGISTRATION NUMBER: 34,858
REFERENCE/DOCKET NUMBER: DCI-033CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-5941
TELEFAX: (617) 227-2700
INFORMATION FOR SEQ ID NO: 2:


```

: SEQUENCE CHARACTERISTICS:
: LENGTH: 4079 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
:
US-08-121-057-2

```

Query Match	2.9%	Score 57	DB 1	Length 4079
Best Local Similarity	46.6%	Pred. No. 3.2e-07		
Matches 257; Conservative	0	Mismatches 285	Indels 9	Gaps 2

QY	1078	tattcattgagcgcgtcccaacatgctgctatcagccaaagtatccaagcttcgcagtgata	1137
Db	2359	TACTTCTTATTATTTGCTCTACCCCTTATCTACCGTGACAGCACTTCCAGGAATCCCACTGA	2416
QY	1138	cggaaaggagtttgggtgcctcgaattcgtcaaaactgctcatcattcaaccggatccaaagca	1197
Db	2419	AGATGGGGTATATGTGCTATAGAAAGTTTGCAACAGCTCTTTGGTTGCTTTTCTATGTGAC	2478
QY	1198	ttataatagaacaataataaactcctaltgtcgaagaactcaagaatcctlttgaagac	1257
Db	2479	TACATCTTTTAAAGGCTTTGTGCCCTCTGTTTCGGAATATCAACAGAGCCCTTTCAGC	2538
QY	1258	gattctctatactatctatgaagaagtggtgaagcttccagttccaaattatatagtgtg	1317
Db	2539	GCTCGTGT---TCTGGCCCTATGTGTGATTAATTAATCCATCTTCCAGGTTGCTGTGATTTCTC	2599
QY	1318	ctctgcacgtctctactgcgtctctccacacttgtttaaacatacttgcaagctctctcgc	1377
Db	2596	TTCCTTACTTTTGTGGCTTTTGGACCTGCTGCTCAAGCCTTTGCTGAGATGTTACGC	2655
QY	1378	ttcggggagatcgtgaattcttacaagaatctggtggaatgcaaaaagtggtggaatctacg	1437
Db	2656	TTTGGTGACAGGATGTTCTATAGAAAGTGTGTGGAATCCACACTCATCCAACTATAT	2715
QY	1438	agaatgtagaatactgcctctgtaataatgatalggttcgaacataat-----acttccg	1491
Db	2716	AGAACTCGGAATGTGTGTCATGACTGCTGCTATATTAATTAAGCTTACAAAGCACTTCTC	2779
QY	1492	tgccttcgcaagaacatataccaagaacatcgcacatataatgcttccctcctagcttcgca	1551
Db	2776	TGGTTTTTTCACAAAGAAATTCAAAATGCTGTCACATGTTAAGCTGTCTTGGCTGTATCTCT	2835
QY	1552	gtcttcacagaactatgcatcgagctccctgtcgtctctctctcctaagaactatggcttttct	1611
Db	2836	GTAGTACACCAAAATATGCTTGGCTGTTTCTGTAGACTTTTCTATCCCGTGTGTCGTG	2895
QY	1612	ggagataatgctt 1622	
Db	2896	CTCTTCATGTT 2906	

RESULT 9
 US-08-509-187D-2
 : Sequence 2, Application US/08509187D
 : Patent No. 5834283
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Chang, Ya-Yuan and Chang, Catherine C.Y.
 : TITLE OF INVENTION: ACYL Coenzyme A: Cholesterol Acyltransferase
 : NUMBER OF SEQUENCES: 9
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: LARIVE & COCKFIELD, LLP
 : STREET: 28 State Street
 : CITY: Boston
 : STATE: Massachusetts
 : COUNTRY: USA
 :
 : ZIP: 02109
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC Compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS

```

1      SOFTWARE: Patentin Release #1.0, Version #1.25
2
3      CURRENT APPLICATION DATA:
4      APPLICATION NUMBER: US/08/509,187D
5      FILING DATE: 31-JUL-1995
6
7      CLASSIFICATION: 435
8
9      PRIOR APPLICATION DATA:
10     APPLICATION NUMBER:
11
12     FILING DATE:
13
14     ATTORNEY/AGENT INFORMATION:
15     NAME: Lamport Hammitte, Ann
16     REGISTRATION NUMBER: 34,858
17     REFERENCE/DOCKET NUMBER: DCI-033cgv
18
19     TELECOMMUNICATION INFORMATION:
20
21     TELEPHONE: (617)227-7400
22
23     TELEFAX: (617)742-4214
24
25     INFORMATION FOR SEQ ID NO: 2:
26     SEQUENCE CHARACTERISTICS:
27     LENGTH: 4079 base pairs
28     TYPE: nucleic acid
29     STRANDEDNESS: single
30     TOPOLOGY: linear
31
32     MOLECULE TYPE: cDNA
33
34     US-08-509-187D-2

```

Query Match	2.9%;	Score 57;	DB 2;	Length 4079;
Best Local Similarity	46.6%;	Pred. No. 3.2e-07;		
Matches 257; Conservative	0;	Mismatches 285;	Indels 9;	Gaps 2

QY	1078	tattcattggtcgtcccaacattggtttacacagcaagatctccacgttcgcagatga	1137
Db	2359	TACTCTATTATTGGTCTCCTACCCCTTATCTACACCGTAGCAGTATCCAGAAATCCACTGTa	2418
QY	1138	cggaaagggtttgggtggtcgtcgaattctggaanaactggtacataaccggaattcggga	1197
Db	2419	AGATGGGGTATGTGCTATGAAGTATGGACAAGGTCTTGTTGGTCTTTTCTATGTGTAC	2478
QY	1198	ttataatagaacaataataataactctatctgtcaggaaactcaagaacatccctttgaaggc	1257
Db	2479	TACTCTCTTAAAGAGCTTGTGTGCCCTTGTTTCGGAATATCAACAGAGAGCCCTTCAGC	2538
QY	1258	gactctctaatatgctattggaagagttctggaagccttcagttcacaattatlatgtygg	1317
Db	2539	GCCTGTCT---TCGTGCTCATATGTGATTTAACTCCATCTTCCACGGGTGTCTATTCTC	2595
QY	1318	ctctgcatagtcttactcgtctctccaccccttgglttaaacataatlgcagagactctctgc	1377
Db	2596	TTCCCTTACTTTTGTGGCTTTTGACACTGCTGGCTCAATAGCCTTTGCTGAGATGTTACGC	2655
QY	1378	ttcggggatctggttaattctcttacaagaatttggatggaatgcataaaagtgtggagattctg	1437
Db	2656	TTTTGGTACAGATGTGTTCAATGAAGTATGTGTGGAACCTCCAGCTCATACTCCAACTATTAT	2715
QY	1438	agaatgtggaatatgctcttcaataatgagatggttcgaactata-----acttccgc	1491
Db	2716	AGAACTCGAATGTGTGTGCTCCATGACTGCTATATTATCTATAGCTTTACAAGAGCTTTCTC	2775
QY	1492	tgccttcgcagcaagaataccaaagaacatcgcacatctatcattgctcttcctagctctgcga	1551
Db	2776	TGGTTTTTCGCCAAGAGATTTCAAAATCTGCTGCCATGTTTAGCTGTCTTGCTGTAATCTGCT	2835
QY	1552	gtcttctcaatgagctatgcatcgcagatccctctgtgctcttcaagaatgtaggtcttctt	1611
Db	2836	GTATGTCACAGAAATATGCTCTGGCTGTCTTGCTTGACACTTTTCTATCCCGTGTCTTGCTG	2895
QY	1612	ggagataatgct 1622	
Db	2896	CTCTTCATGTT 2906	

RESULT 10
US-09-121-396-2
; Sequence 2, Application US/09121396

[illegible][illegible]

```

RESULT 15
US-09-007-005-32/C
: Sequence 32, Application US/09007005B
: Patent No. 6258558
: GENERAL INFORMATION:
: APPLICANT: Szostak, Jack W.
: APPLICANT: Roberts, Richard W.
: APPLICANT: Liu, Rihue
: TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
: TITLE OF INVENTION: FUSIONS
: FILE REFERENCE: 00786/350003
: CURRENT APPLICATION NUMBER: US/09/007, 005B
: CURRENT FILING DATE: 1998-01-14
: EARLIER APPLICATION NUMBER: 60/035,963
: EARLIER FILING DATE: 1997-01-27
: EARLIER APPLICATION NUMBER: 60/064,491
: EARLIER FILING DATE: 1997-11-06
: NUMBER OF SEQ ID NOS: 33
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 32
: LENGTH: 248
: TYPE: RNA
: ORGANISM: Homo sapiens
US-09-007-005-32

```

Query Match	1.8%;	Score 35.6;	DB 4;	Length 248;
Best Local Similarity	22.5%;	Pred. NO. 0.23;		
Matches 53;	Conservative 77;	Mismatches 106;	Indels 0;	Gaps 0;

